Perspectives on contexts

Paolo Bouquet, Luciano Serafini and Rich Thomason (Eds.)

Contents

Introduction: Perspectives on Context

PAOLO BOUQUET, LUCIANO SERAFINI, RICHMOND THOMASON

You are standing by a stream, facing your friend across the running water. Your friend wants to know which way to go to find the bridge, which is to your left. You say, "Go to the right." Your friend turns to the right and finds the bridge. If you had said "Go to the left," your friend would have gone the wrong way, and she would have been right to blame you for the error—normal adults are able to recognize such perspectival differences, and to take them into account appropriately in conversation. The contextualized version that you start with in this case, "The bridge is to the left," is the representation that is most closely connected to my own perceptions and actions. If I can see the bridge, I see that it is to my left. If I know that it is to my left, I know which way I should go to get to it. (If, for instance, I only know is that the bridge is to the north, I may not know which way to go.) For the same reason, it is easier for you to act on my instructions if they are packaged in a way that relates them immediately to the actions you need to perform. But if I make an error in repackaging the message, or you think that I am speaking from my perspective, the intended meaning will be reversed.

Often, our reasoning about such contextual matters is automatic and unconscious. But it can be conscious and reflective. In deciding whether to tell you to turn north or turn right when I am giving you driving directions to my house, I may have to ask myself whether I'm sure which

way you'll be driving, and whether you will know which way is north. In the bridge example, the speaker has to be aware of the perspectival dependence of words like 'left' and 'right', to temporarily adopt the perspective of the addressee, and to recontextualize the information from this perspective. Some reasoning tasks—for instance, matching the information on a map to what we are seeing—require the reasoner to move smoothly between decontextualized and contextualized forms of the same information.

Contextualization makes thought and communication practicable by allowing us to leave out the extra baggage that is needed in order to make representations interpretable at other times, in other places, and to other audiences. But the contextualization that makes our thought and language more vivid and compact can be a challenge to observers at a distance who are trying to interpret our thoughts or replicate our reasoning. This interpretive challenge has become particularly acute as we encounter the problem of using information from different sources (documents, databases, records in various formats) that have been prepared locally, perhaps for a limited audience, but now are made available to the world at large by modern technology. The ability to process contextualized information smoothly is a hallmark of intelligent commonsense reasoning, but like many such reasoning abilities, replicating it by robust automated reasoning presents a serious challenge for research in Artificial Intelligence.

But the contextualization of language has preoccupied philosophers for much longer than this. Anyone who has tried to teach Frege's ideas about meaning and language is likely to be struck by the fact that, while the views of "On sense and reference" Frege (1960) can at least be summarized in a fairly straightforward way, it is much more difficult to do this for the things that Frege had to say in "The thought" Frege (1956) about contextualized language. Important subsequent programs in the philosophy of language and semantics have more or less ignored problems of contextuality, hoping perhaps that the things that can be said about the simplest sorts of language will apply to the entire range of contextual phenomena, and that contextualization will require no theoretical insights that are fundamentally new. (The programs association with Rudolf Carnap and Richard Montague have this flavor.) But the more we discover about language use, and the more we attend to realistic examples, the more difficult it is to maintain such hopes.

At the same time that context has come to be seen as more central and more problematic for the philosophy of language, it has also become more important as a potential tool for core philosophers. Following David Lewis' suggestions in "Scorekeeping in a language game,"

Lewis (1979) many subsequent philosophers have explored the idea that context might help to disarm philosophical skepticism. Other philosophers have used it to obtain new insights into fundamental epistemplogical problems. See Preyer and Peter (2005) for more information on this topic.

Context has a long history also in different areas of artificial intelligence. Perhaps the first reference can be traced back to Richard Weyhrauch and his work on mechanizing logical theories in an interactive theorem prover called FOL, Weyhrauch (1979). However, it became a widely discussed issue only in the late 1980s, when John McCarthy proposed the formalization of context as a crucial step toward the solution of the problem of generality.

> When we take the logic approach to AI, lack of generality shows up in that the axioms we devise to express common sense knowledge are too restricted in their applicability for a general common sense database .... Whenever we write an axiom, a critic can say that the axiom is true only in a certain context. With a little ingenuity the critic can usually devise a more general context in which the precise form of the axiom doesn't hold." McCarthy (1987).

For this reason, McCarthy sees an explicit theory of context as an essential part of his program of formalizing common sense knowledge and reasoning. It enables axiomatizations to be extended and improved by contextualizing localizing axioms, rather than having to remove them and replace them with other axioms.

An independent approach, motivated by the intuition that common sense reasoning is always "contextual" (namely local to a subset of an agent's knowledge base, and the exported to other contexts if and when required) was proposed by Fausto Giunchiglia in work published in 1993 Giunchiglia (1993), and since then developed in different directions. These two approaches were conceptually and technically compared in Serafini and Bouquet (2004).

More recently, contexts have been seen as providing crucial support for the Semantic Web; see, for instance, Bouquet et al. (2003), Guha et al. (2004). In developing the Semantic Web the impossbility has been widely recognized of working with a unique ontology (repository of taxonimically organized knowledge). This has led to a number of proposals for ontology modularization and cross-ontology mappings. Here, theories of context are finding a direct area of application.

These developments serve in part to account for the recent interest in context across a number of disciplines, and for the success of the biennial interdisciplinary conferences on modeling and using context that have been held regularly since 1997.

The papers in this volume, collected from one of these conferences, reflect the range and sophistication of current thinking on the topic of context. They clearly show how insights from all the areas of cognitive science can be combined to improve our understanding of this important topic.

In "On the dimensions of context dependence," Massimo Benerecetti, Paolo Bouquet, and Chiara Ghidini address a theme that is addressed in several other papers in this context: the use of symbolic logic to formalize contextual reasoning. By examining the forms of contextual reasoning that have been studied in Artificial Intelligence, they propose that these forms fall into three general categories. In the course of motivating and defending this classification, they present and analyze a number of examples of reasoning that involve context.

Chiara Ghidini and Fausto Giunchiglia provide an overview in "What is Local Models Semantics?" to the approache to the logical formalization of context, due originally to Giunchiglia (1993). They illustrate the approach with an extended example, showing how it provides a way of representing "local" or contextualized theories while at the same time providing a model-theoretic semantics that can be used to validate translations among the theories.

"Contextual Intensional Logic: Type-Theoretic and Dynamic Considerations," by Richmond Thomason, describes work reported in a series of papers dating from 1997, which develops an approach to the logic of context based on higher-order logic. This paper concentrates on a dynamic version of the logic, which can be used to formalize reasoning of the sort described in Chapter 1 of this volume, and in particular operations that allow a reasoning agent to "move" from one context to another.

With "The search for the semantic grail," John Perry turns to issues in the philosophy of language, concentrating on the concept of semantic content. Like many contemporary philosophers of language, Perry thinks of semantic content as something that is associated with thoughts and linguistic expressions by a context: the content of a sentence like 'I was in Sacramento yesterday' will depend, in general, on the context with which it is associated. This picture of things, which stresses the role of context as a means of associating contents with "indexical expressions," was articulated by David Kaplan and presents philosophers with the problem of clarifying the associated idea of semantic content. Perry argues that there is no single such concept serving the fundamental purposes of philosophical semantics—instead, there is a related family of different sorts of semantic contents.

In "On a Proposal of Strawson Concerning Context versus 'What Is

Said','" Varol Akman discusses another approach to the philosophical problem of content, inspired by the Oxford philosopher Peter Strawson. Like Perry, Strawson distinguishes several notions of content—but his classification is motivated by different considerations, and cuts in different ways. Unlike Perry, who concentrates on a relatively narrow and closely related set of problems having to do with the relation of referential content to the perspective of an agent, Strawson's classification raises the issue of how the meaning of an utterance in many cases can be disassociated from the linguistic vehicle, and associated with the intentions of the speaker. This notion of "pragmatic content" owes much to the work of Paul Grice, and suggests an approach to content that is much more social and open-ended than that of the philosophers concerned with semantic content.

Issues concerning pragmatic content or "speaker meaning" are unavoidable when one considers the interpretation of extended discourses, and especially of conversations, rather than the assignment of contents-in-context to isolated sentences. During much of the twentieth century, logically-minded philosophers tended to concentrate on the latter problem, whereas a more informal group of philosophers—some inspired by the later Wittgenstein, and some by Oxford ordinary language philosophy—concerned themselves with more naturally occurring linguistic data and extended discourse. More recently, developments in logic and linguistics (and especially, the emergence of discourse representation theory in linguistics and of dynamic logic) have provided a framework in which logical techniques can be fruitfully applied to extended discourse. In "Epistemological foundations for the representation of discourse context," Horacio Arló-Costa shows how these techniques can shed light not only on the nature of context as it figures in the interpretation of discourse, but in its rule-governed evolution in the course of a conversation.

Taking a different turn, in "Context and logical form" Wolfram Hinzen considers the foundations of context as part of a linguistic theory, and—more generally—as a component of cognitive science. Taking a view of the nature of linguistics that appears to be influenced by Noam Chomsky, and that is broadly compatible with Chomsky's recent views, Hinzen asks what place context can have within a properly scientific—that is, a naturalistic—account of the human language faculty. Hinzen argues that the role context can play in such a theory is actually quite restricted, and that it consists primarily of the presence of contextual variables in the forms that are manipulated by a cognitive language-processing module, acting more or less independently of general-purpose common sense reasoning. On the negative side, Hinzen

urges that theories that depend heavily on context, such as dynamic accounts of linguistic meaning, are too open-ended and normative to provide a proper basis for scientific investigations. On the positive side, he refers us to suggestions of Juan Uriagerika's concerning the role of contextual variables in Logical Form.

If we accept something like Hinzen's division between semantics and pragmatics, and even embrace the conclusion that pragmatics cannot be a part of a naturalistic science, it seems to contain a great deal of what is most interesting and important about language, and perhaps one can develop a research program that can address these matters, even if the program cannot be part of naturalistic science. In "Truth-conditional pragmatics: an overview," François Recanati describes one such approach. According to truth-conditional pragmatics, the role of norms and of general common-sense reasoning is more radical in the determination of semantic content than many philosophers have assumed. Recanati's paper makes the case for this position, and explores some of the issues that arise within this approach to meaning.

In "Context and contract," Carlo Penco continues the exploration of pragmatic theory, this time from a more philosophical perspective. Penco distinguishes different notions of context that have been invoked by philosophers of language; the resulting classification is motivated in part by the considerations in play in Perry's and Akman's classifications of content. Against Christopher Gauker's criticism of the value of the notion of a shared content in communication, Penco points to cases in which conversants actually seem to be explicitly negotiating a meaning.

The last two articles in this volume turn to applications of ideas about context of the sort that were discussed in the previous example: in aesthetics and in the philosophy of science.

In "Fictional Contexts," Andrea Bonomi uses contexts to distinguish the reference of characters as such (where the character is a set of properties and relations), and characters in the context of the fictional work in which they appear, where they function as actual persons. Bonomi goes on to show how techniques from semantics can be used to resolve otherwise puzzling sentences in which reference is made to fictional characters and objects. Bonomi makes a good case for the value of logical ideas about context in intepreting fictional discourse and discourse about fiction.

The role of the intellectual setting in which a scientist works in scientific discovery and the evaluation of scientific theories, and even in the assignment of content to apparently decontextualized scientific terms, emerged as a fundamental issue in the development of philosophy of science during the twentieth century. In his "Context and Philosophy

of Science" Robert Young seeks to characterize a notion of context that is appropriate for scientific theories in useful in understanding the issues concerning contexualization in the philosophy of science. This paper shows that formal ideas from the theory of context can illuminate challenging issues in this area of philosophy.

Ranging from applications in Artificial Intelligence to the foundations of semantics and pragmatics, this volume provides an excellent introduction to contemporary thinking about context, and illustrates ways in which these ideas can advance our understanding of issues in philosophy.

List of contributors:

Varol Akman: Professor of Computer Engineering and Chair, Department of Philosophy Bilkent University.

Horacio Arló-Costa: Associate Professor, Philosophy Department, Carnegie Mellon University

Massimo Benerecetti: Associate Professor in Computer Science at the Università di Napoli "Federico II"

Andrea Bonomi: Professor of Semantics of Natural Languages, Dipartimento di Filosofia, Universitá degli Studi di Milano

Paolo Bouquet: Associate Professor, Dipartimento di Informatica e Telecomunicazioni, Universitá degli Studi di Trento

Chiara Ghidini:: Research Scientist, Istituto per la Ricerca Scientifica e Tecnologica, Istituto Trentino di Cultura

Fausto Giunchiglia: Head of the Department of Information and Communication Technology, Universitá degli Studi di Trento

Wolfram Hinzen: Senior Lecturer, Faculteit der Geesteswetenschappen, Universiteit van Amsterdam

John Perry: Professor, Philosophy Department, Stanford University

Carlo Penco: Associate Professor of Philosophy of Language Department of Philosophy, Universit di Genova

François Recanati: Research Director, Centre National de la Recherche Scientifique, Paris

Luciano Serafini: Senior Researcher, Istituto per la Ricerca Scientifica e Tecnologica, Istituto Trentino di Cultura

Richmond H. Thomason: Professor of Philosophy, Univeristy of Michigan

Roger A. Young: Lecturer, Philosophy Department, University of Dundee

References

Bouquet, Paolo, Fausto Giunchiglia, Frank van Harmelen, Luciano Seranifini, and Heiner Stuckenschmidt. 2003. C-OWL: Contextualizing ontologies. In D. Fensel, K. Sycara, and J. Mylopoulos, eds., *The Semantic Web— ISWC 2003: Second International Semantic Web Conference*, pages 164–179. Berlin: Springer-Verlag.

Frege, Gottlob. 1956. The thought. *Mind, New Series* 65(259):289–311. Originally published in 1918 as "Die Gedanke." Translated by P.T. Geach.

Frege, Gottlob. 1960. On sense and reference. In P. Geach and M. Black, eds., *Translations from the Philosophical Writings of Gottlob Frege*, pages 56–78. Oxford: Basil Blackwell. Originally published in 1892 as "Über Sinn und Bedeutung." Translated by Max Black.

Giunchiglia, Fausto. 1993. Contextual reasoning. *Epistemologica, Special Issue on I Linguaggi e le Macchine* 16:345–364.

Guha, Ramanathan V., Rob McCool, and Richard Fikes. 2004. Contexts for the semantic web. In S. A. McIlraith, D. Plexousakis, and F. van Harmelen, eds., *The Semantic Web—ISWC 2004: Third International Semantic Web Conference*, pages 32–46. Berlin: Springer-Verlag.

Lewis, David K. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8(3):339–359.

McCarthy, J. 1987. Generality in Artificial Intelligence. *Communications of ACM* 30(12):1030–1035. Also in V. Lifschitz (ed.), *Formalizing common sense: papers by John McCarthy*, Ablex Publ., 1990, pp. 226–236.

Preyer, Gerhard and Georg Peter, eds. 2005. *Contextualism in Philosophy*. Oxford: Oxford University Press. ISBN 0199267405 (hbk.), 0199267413 (pbk.).

Serafini, Luciano and Paolo Bouquet. 2004. Comparing formal theories of context in AI. *Artificial Intelligence* 155(1–2):41–67.

Weyhrauch, Richard W. 1979. Prolegomena to a theory of mechanized formal reasoning. *Artificial Intelligence* 13(1–2):133–170.

1

On the dimensions of context dependence

Massimo Benerecetti, Paolo Bouquet, Chiara Ghidini

In this chapter we propose to re-read the past work on formalizing context as the search for a logic of the relationships between partial, approximate, and perspectival theories of the world. The idea is the following. We start from a very abstract analysis of a context dependent representation into three basic elements. We briefly show that all the mechanisms of contextual reasoning that have been studied in the past fall into three abstract forms: *expand/contract*, *push/pop*, and *shifting*. Moreover we argue that each of the three forms of reasoning actually captures an operation on a different dimension of variation of a context dependent representation, *partiality*, *approximation*, and *perspective*. We show how these ideas are formalized in the framework of MultiContext Systems, and briefly illustrate some applications.

1.1 Introduction

In the last twenty years, the notion of context has become more and more central in theories of knowledge representation in Artificial Intelligence (AI). The interest in context is not limited to AI, though. It is discussed and used in various disciplines that are concerned with a theory of representation, such as philosophy, cognitive psychology, pragmatics, linguistics. Despite this large amount of work, we must confess that we are very far from a general and unifying theory of context. Even if we restrict the focus to theories of representation and

language, it is very difficult to see the relationship between different works on contextual reasoning. As an example, there are good pieces of work on utterance contexts, belief (and other intensional) contexts, problem solving contexts, and so on, but it is not clear whether they address different aspects of the same problem, or different problems with the same name.

In this chapter we propose to re-read the past work on context as the search for a logic of the relationships between partial, approximate, and perspectival theories of the world. The idea is the following. We start from an very abstract analysis of a context dependent representation into three basic elements: a collection of parameters (the contextual dependencies), a value for each parameter, and a collection of linguistic expressions (the explicit representation). Then, we briefly show that all the mechanisms of contextual reasoning that have been studied in the past fall into three abstract forms, *expand/contract*, *push/pop*, and *shifting*, each corresponding to an operation on one of the basic elements of the representation. Then, we argue that each of the three forms of reasoning actually captures an operation on a different dimension of variation of a context dependent representation, *partiality*, *approximation*, and *perspective*. This leads us to the conclusion that, at a suitable level of abstraction, *a logic of contextual reasoning is precisely a logic of the relationships between partial, approximate, and perspectival theories of the world*. We show how these ideas are formalized in the framework of MultiContext systems, and briefly illustrate some applications.

1.2 Contexts as boxes

In general, a representation is called context dependent when its content cannot be established by simply composing the content of its parts. In addition, one has to consider extra information that is left implicit in the representation itself. In Giunchiglia and Bouquet (1997), this notion of a context dependent representation is illustrated by introducing the so-called metaphor of the box (figure 1). A context dependent representation has three basic elements: a collection of parameters $P_1, \ldots, P_n, \ldots$, a value V_i for each parameter P_i, and a collection of linguistic expressions that provide an explicit representation of a state of affairs or a domain. The intuition is that the content of what is inside the box depends (at least partially, and in a sense to be defined) upon the values of the parameters associated with box. For example, in a context in which the speaker is John (i.e. the value of the parameter 'speaker' is set to John), the content (the intension, if you prefer) of the pronoun 'I' will be John, but this is not the case in a context in

which the speaker is Mary.

P1=V1 ..... Pn=Vn

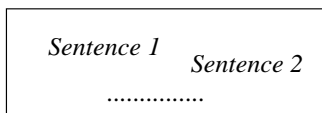*Sentence 1*
  Sentence 2
        ....

FIGURE 1 Contexts as boxes.

Starting from the metaphor of the box, it is quite easy to see that a theory of contextual reasoning is faced with a number of philosophical problems. A partial list includes: What features of context should be included among the parameters? Is it possible to specify *all* the relevant parameters, or the collection is always incomplete? How is the representation affected when the collection of parameters or their values changes? Can we get rid of parameters and get a context independent representation of the contents of a box? What is the relationship between the parameters of different boxes? How does this relationship affect the relationship between the contents of different boxes?

Since the goal of this chapter is not to provide a general foundation for a theory of context, we will not propose an answer to the issues above. Indeed, the analysis of the patterns of contextual reasoning is meant to hold no matter what solutions one adopts to these fundamental issues.

1.3   Forms of contextual reasoning

Mechanisms for contextual reasoning have been studied in different disciplines, though with different goals. A very partial list includes:*reflection* and *metareasoning* Weyhrauch (1980), Giunchiglia and Serafini (1994), *entering and exiting context, lifting, transcending context* Guha (1991), McCarthy (1993), Buvac and Mason (1993), *local reasoning, switch context* Giunchiglia (1993), Bouquet and Giunchiglia (1995), *parochial reasoning* and *context climbing* Dinsmore (1991), *changing viewpoint* Attardi and Simi (1995), *focused reasoning* Laird et al. (1987)). As a matter of fact, it is very difficult to see the relationship between these different works. We try to put some order in this situation by addressing the problem of identifying the general patterns of contextual reasoning, namely the general mechanisms that people use to reason with information (i) whose representation depend on a collection of contextual parameters, and (ii) which is scattered across a multiplicity of different contexts.

Our proposal is that all the forms of contextual reasoning that are discussed in the literature fall into three basic patterns, according to the element of the box that they affect: the representation, the collection of parameters, and the parameters' values.

Expand/Contract. A first general form of contextual reasoning (depicted in Figure 2) is based on the intuition that the explicit representation associated with a specific context does not contain all the facts potentially available to a reasoner, but only a subset of them. As a consequence, depending on the circumstances, the subset which is explicitly taken into account can be expanded (typically because some new input from the external environment makes it necessary to consider a larger collection of facts), or contracted (typically because the reasoner realizes that some facts are not relevant on a given occasion). An example of expansion is the Glasgow-London-Moscow (GLM) example McCarthy (1991), Bouquet and Giunchiglia (1995): when reasoning about traveling from Glasgow to Moscow via London, we normally do not include in the problem solving context the precondition that one must be dressed to get on a plane; however, if one's clothes are stolen at London airport, being clothed becomes a relevant precondition for the success of the travel plan, and therefore the original problem solving context must be expanded with facts about social conventions and buying clothes.
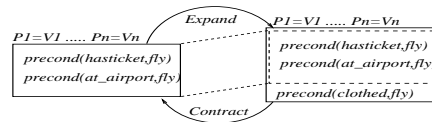

FIGURE 2 Expand/Contract.

In general, expansion and contraction are used to adjust a particular representation to a problem or to a given goal. The way problem solving contexts are built in CYC (using the strategy of lift-and-solve Guha (1991)), the mechanism of building appropriate mental spaces Fauconnier (1985) or partitioned representations Dinsmore (1991), and the process of selecting the relevant facts to interpret an utterance Sperber and Wilson (1986) are typical examples of this pattern of contextual reasoning.

Push/Pop. The content of a context dependent representation is partly encoded in the parameters outside the box, and partly in the sentences inside the box. Some authors propose reasoning mechanisms

for altering the balance between what is explicitly encoded inside the box and what is left implicit (i.e. encoded in the parameters). Intuitively, the idea is that we can move information from the collection of parameters outside the box to the representation inside the box, and vice versa. We call these two mechanisms *push* and *pop* to suggest a partial analogy with the operations of adding (pushing) and extracting (popping) elements from a stack. In one direction, *push* adds a contextual parameter to the collection outside the box and produces a flow of information from the inside to the outside of the box, that is part of what was explicitly encoded in the representation is encoded in some parameter. In the opposite direction, *pop* removes a contextual parameter from the collection outside the box and produces a flow of information from the outside to the inside, that is the information that was encoded in a parameter is now explicitly represented inside the box.
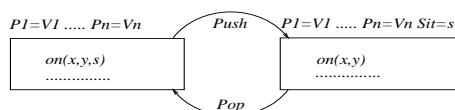

FIGURE 3 Push/Pop.

Consider, for instance, the well known AboveTheory scenario, introduced in McCarthy (1993). The fact that block x is on block y in a situation s is represented as $on(x, y, s)$ in a context c with no parameter for situations. This is because in some cases we want to leave implicit the dependence on the situation s (typically, when we don't want to take situations into account in reasoning). This means that the situation can be encoded as a parameter, and the representation can be simplified to $on(x, y)$. Push is the reasoning mechanism which allows us to move from $on(x, y, s)$ to $on(x, y)$ (left-to-right arrow in figure 3), whereas pop is the reasoning mechanism which allows us to move back to $on(x, y, s)$ (right-to-left arrow in figure 3). Hence, push and pop capture the interplay between the collection of parameters outside the box and the representation inside the box.

It is worth noting that the mechanism of entering and exiting context proposed by McCarthy and others can be viewed as an instance of push and pop. Suppose we start with a sentence such as $c_0 c : p$, whose intuitive meaning is that in context c_0 it is true that in context c the proposition p is true. The context sequence $c_0 c$ can be viewed as the reification of a collection of parameters. Exiting c pops the context sequence, and the result is the formula $c_0 : ist(c, p)$, where the dependence

on c is made explicit in the representation $ist(c, p)$ ($ist(c, p)$ is the main formula of McCarthy's formalism, asserting that a p is true in context c); conversely, entering c pushes the context sequence and results in the formula $c_0 c : p$, making the dependence on c implicit in the context sequence. Other examples of push/pop are: *reflection up* to pop the collection of parameters and *reflection down* to push it in Giunchiglia and Serafini (1994); the rule of *context climbing* to pop the collection of parameters, and the rule of *space initialization* to push it in Dinsmore (1991).

Shifting. Shifting changes the value of one or more contextual parameter, without changing the collection of parameters. The name 'shifting' is inspired to the concept of shifting in Lewis (1980). The intuition is that changing the value of the parameters shifts the interpretation of what is represented inside the box.
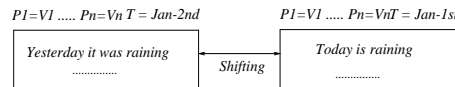

FIGURE 4 Shifting.

The simplest illustration of shifting is again indexical expressions. The fact that on January 1st it is raining is represented as 'Today is raining' in a context in which time is set to January 1st, but it is represented as 'Yesterday it was raining' if the value of time changes to January 2nd. As it is shown in figure 4, *shifting* is the reasoning mechanism which allows us to move from one representation to the other by changing the value of the parameter time, provided we know the relationship between the two parameter's values. Another very common example of shifting is when the viewpoint changes, e.g. when two people look at the same room from opposite sides (what is right for the first will be left for the other). A third case is categorization . For the supporters of team A, the members and the supporters of team B are opponents, and vice versa for the supporters of team B. And the examples can be multiplied.

In the literature, we can find different instances of shifting. Kaplan's notion of *character* is the semantical counterpart of this reasoning mechanism with indexical languages; Guha and McCarthy formalize a form of shifting using the notion of *lifting* Guha (1991); Dinsmore introduces the notion of *secondary context*.

1.4   Dimensions of context dependence

Our next step is to show that the three forms of contextual reasoning actually operate each on a fundamental dimensions of a context dependent representation: *partiality*, *approximation*, and *perspective*. We start with a more precise characterization of partiality, approximation, and perspective.

Partiality. We say that *a representation is partial when it describes only a subset of a more comprehensive state of affairs.* We observe that the notion of partiality can be analyzed from two different perspectives: metaphysically, a theory is partial if it does not cover the entire universe; however, cognitively, a representation is partial if it does not cover the totality of what an agent can talk about. For our present purposes, either perspective is acceptable, even though our general attitude is in favor of the cognitive view.

Perhaps the more intuitive example of partial theories are domain specific theories. For instance, a theory about the Italian cuisine is partial because it does not provide information about Indian or French cuisine, about soccer, about quantum mechanics. A different usage of partial theories is in problem solving. Given a problem, people seem to be capable of circumscribing what knowledge is relevant to solve it, and disregard the rest. In this case, assumptions on what is relevant act as contextual parameters. Partial theories are also used in theories of linguistic communication. When a speaker says something to a hearer, it is assumed that the latter interprets what the speaker said in some context. According to Sperber and Wilson (1986), '[a] context is a psychological construct, a subset of the hearer's assumptions about the world'. Such a context includes the set of facts that the hearer takes to be relevant in order to assign the correct interpretation to what the speaker said. In this sense, it is a partial theory.

Partiality is a relative notion. Intuitively, there is a partial order between partial representations. Therefore a representation can be more or less partial of another one. Two partial representations may also overlap. We do not further discuss these aspect here. We only need to make clear the idea that partiality is a dimension along which a representation may vary.

Approximation. We say that a *a representation is approximate when it abstracts away some aspects of a given state of affairs.* A representation of the blocks world in terms of the binary predicates $on(x, y)$ e $above(x, y)$ is approximate, because the time (situation) is abstracted away.

As for partiality, approximation is a relative notion: a representation

is approximate because it abstracts away details that another representation takes into account. The representation $on(x, y)$ and $above(x, y)$ is more approximate than the representation $on(x, y, s)$ and $above(x, y, s)$ because the first abstracts away the dependence on the situation. Of course, an open point is whether there is such a thing as a non approximate representation of a state of affairs. This would be a sort of least approximate representation, namely a representation which is less approximate than anyone else. We avoid committing to one position or the other; here we are interested in the reasoning mechanisms that allow us to switch from a more to a less approximate representation (and vice versa), and not in the epistemological status of representations.

Perspective. A third dimension along which a representation may vary is perspective. We say that *a representation is perspectival when it encodes a spatio-temporal, logical, or cognitive point of view on a state of affairs.*

The paradigmatic case of spatio-temporal perspective is a given by indexical languages. A sentences such as 'It's raining (here)(now)' is a perspectival representation because it encodes a spatial perspective (i.e. the location at which the sentences are used, the speaker's current 'here') and a temporal perspective (i.e. the time at which the sentences are used, the speaker's current 'now'). The philosophical tradition shows us that some sentences (e.g. 'Ice floats on water') encode a logical perspective as well, because they implicitly refer to 'this' world, namely the world in which the 'here' and 'now' of the speaker belong (the same sentence, if uttered in a world different from our world, might well be false). Thus Kaplan includes a world among the features that define a context, and uses this world to interpret the propositional operator 'actually'.

Indexicals are not the only type of expressions that encode a physical perspective. Suppose, for example, that two agents look at the same object (for example the magic box of figure 5). Because of their different viewpoints, the representation of what they see is completely different, and the same ball can be described as being on the right by **Side** and as being on the left by **Front**.

A subtler form of perspective is what we call cognitive perspective. It has to do with the fact that many representation encode a point of view which includes a collection of beliefs, intentions, goals, and so on. Cognitive perspective is very important in the analysis of what is generally called an *intensional context*, such as a belief context. John and Mary may have dramatically different beliefs about Scottish climate, even if they represent the same universe of discourse (or portion of the

world) at the same level of approximation. We don't see any other way of making sense of this difference than that of accepting the existence of a cognitive perspective, which is part of what determines the context of a representation.

At this point, we are ready to justify our claim that the three forms of contextual reasoning are precisely mechanisms that operate on the dimensions of partiality, approximations, and perspective:

- Expand/Contract is the reasoning mechanism that allows us to vary the degree of partiality by varying the amount of knowledge which is used in the representations of the world.
- Push/Pop is the reasoning mechanism that allows us to vary the degree of approximation by regulating the interplay between the collection of parameters outside and the explicit representation inside a box.
- Shifting is the reasoning mechanism that allows us to change the perspective by taking into account the 'translation' of a representation into another when the value of some contextual parameter is changed.

As a consequence of our claim, a logic of contextual reasoning must formalize the reasoning mechanisms of expand/contract push/pop, and shifting and use them to represent the relationship between partial, approximate, and perspectival representations. Our final step is to show that MultiContext systems satisfy this requirement, and to validate this assertion by analyzing in detail some applications of MultiContext systems.

1.5 A logic of contextual reasoning: MultiContext Systems

In the past, various logics have been proposed which formalize one aspect or the other of such a logic of contextual reasoning. For example, Kaplan's logic of demonstratives is a logic which allows only for a multiplicity of perspectival representations (partiality and approximation are left unchanged from one context to the other) and, consequently, provides only mechanisms for shifting (in the form of the semantic notion of character). Buvač and Mason's propositional logic of context allows for a multiplicity of partial, approximate, and perspectival representations, and provides the machinery for expand/contract, push/pop, and shifting; however, it formalizes a quite weak form of partiality (via the use of partial functions for interpreting a global language) and only a special form of push/pop (i.e. making explicit or implicit the context itself).

MultiContext systems (MCS) Giunchiglia and Serafini (1994), and their *Local Models Semantics* (LMS) Ghidini and Giunchiglia (2001), provide a logic for contextual reasoning based on the principles of *locality* and *compatibility*. These principles are stated and discussed in Chapter 2, where LMS is introduced and explained in detail. For the sake of our presentationdiscussion, we rewrite these principles as follows:

1. each context c_i is associated with a different formal language L_i, used to describe what is true in that context. The semantics of L_i is *local* to the context itself. Therefore each context has its own set of local models M_i, and local satisfiability relation $\models_i$;
2. the relations between different contexts are modeled by means of *compatibility* relations between (sets of) local models of the different contexts.

We believe that principle of locality and principle of compatibility provide LMS and MCS with the capability of being a suitable logic of the relation between partial, approximate, and perspectival representations. For lack of space, we focus the discussion of our claim on LMS. A similar analysis applies to the axiomatic system MCS.

By associating distinct languages and local semantics to different contexts, LMS allows for different partial, approximate, and perspectival representations. The most intuitive case is partial representations. Simple examples are: the language might contain only a subset of a more comprehensive set of symbols, the class of models might satisfy only a subset of a more comprehensive set of axioms, or rules of wellformedness. Second, approximate representations. A simple case is the AboveTheory example: a context might contain the binary predicate $on(x, y)$ or the ternary predicate $on(x, y, s)$ depending on the fact that the it abstracts away or represents the dependence on the situation. Third, perspectival representations. An example is the fact that the truth value of a formula in a context might depend on the perspective from which the world is represented. The truth value of the formula might change in different contexts, depending on the corresponding shift of perspective.

By modeling compatibility relations between different contexts as relations among the (sets of) local models of the different contexts LMS allows one to represent the relations existing between a multiplicity of partial, approximate, and perspectival representations of the world. For instance, if the relation contains a pair $\langle models(c_1), models(c_2) \rangle$ composed by a set $models(c_1)$ of models of context c_1 and a set $models(c_2)$ of models of context c_2, and all the models in $models(c_1)$ are obtained
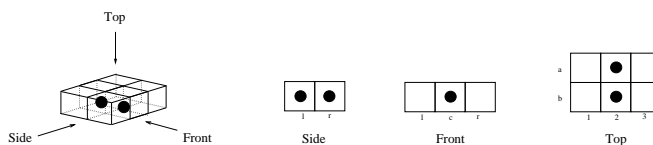
FIGURE 5 The magic box and its partial views.

as the expansion of a model in $models(c_2)$, then it is easy to observe that c_2 describes a portion of the world which is a subset of the portion described by c_1. By studying and classifying the different relations existing among the (sets of) local models of the different contexts we might, in principle, try to classify the many different relations existing among different partial, approximate, and perspectival representations. Unfortunately, even if we restrict ourselves to model each context c_i by mean of a first order language and the classical semantics, we must admit that we are still far from having a (even partial) classification of these many different relations. Although some of them are very easy to identify, as the relation of expansion mentioned above, relations between partial, approximate and perspectival representations may be, in general, much trickier. Nonetheless, by analyzing existing applications of LMS and MCS we are able to show that LMS and MCS have been used to represent context-based representation and reasoning in terms of the relations among partial, approximate, and perspectival representations. In the rest of the section briefly show the result of our analysis. This provides a first evidence of the fact that LMS is a logic of the relations between partial, approximate and perspectival representations. This provides also a fist motivation for a future work on studying and classifying the many different relations existing among different partial, approximate, and perspectival representations.

Viewpoints. A paradigmatic example of reasoning with viewpoint is the Magic Box (MB) example, developed in Benerecetti et al. (2000).

> There are three observers, **Top**, **Side**, and **Front**, each having a partial view of a box as shown in the top part of Figure 5. **Top** sees the box from the top, and **Side** and **Front** see the box from two different sides. The box consists of six sectors, each sector possibly containing a ball. The box is "magic" and **Side** and **Front** cannot distinguish the depth inside it. The bottom part of Figure 5 shows the views of the three agents corresponding to the scenario depicted in the top part. **Top**, **Side**, and **Front** decide to test their new computer program ϵ by submitting the following puzzle to it. **Side** and **Front** tell ϵ their partial views. Then they ask ϵ to guess **Top**'s view of the box.

Benerecetti et al. (2000) describes a formalization of the reasoning process of ϵ in solving the puzzle, by mean of the four contexts depicted in figure 6. Contexts **Side** and **Front** contain the program's
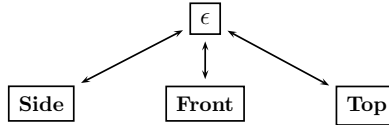

FIGURE 6 The four contexts in the MB example.

representation of **Side**'s and **Front**'s knowledge; context **Top** contains the program's representation of **Top**'s knowledge, and is the context in which it will try to build the solution; context ϵ contains the knowledge that the computer program has about the game, namely what the relations among the other contexts are.

According to our classification of dimensions of a context dependent representation, the representations of the different contexts **Side** and **Front**, **Top**, and ϵ may vary along three dimensions: partiality, approximation, and perspective. Focusing on partiality, the different contexts are related to different specific domains. For instance, **Side** can only talk about the (non) presence of a ball in the left or right sector it sees, **Front** can talk about the (non) presence of a ball in the left, or the central or right sector it sees, **Top** can talk about the presence of a ball in each one of the six sectors, while ϵ needs only to talk about how the pieces of knowledge contained in each one of the contexts above are related to each other. Focusing on approximation, we notice that the description of (a portion of) the world in **Side**, **Front**, and **Top** is given in terms of balls and sectors of the box, whereas the description in context ϵ concerns how to relate the information coming from the different observers. In order to do this, context ϵ needs to make explicit some information that was implicit in the observers' contexts. In particular, it needs to make explicit what information comes from what observer. This is an example of push/pop and is related to the different levels of approximation of the different contexts. In this case we say that the representation in **Side**, **Front**, and **Top** is more approximate than the one in ϵ, because the first ones abstract away what information comes from what observer. Focusing on perspective, each of the observer's contexts expresses knowledge about the box which depends on the observer's physical perspective. For example, the fact that **Side** sees a ball in the left sector (from his point of view) is different from **Front** seeing a ball in the left sector (from his point of view). Since their

perspectives are different, the same description (e.g., 'A ball is in the left sector') may, thus, have a different meaning in different contexts.

Formally, the specific domains of **Side**, **Front**, and **Top** are described by three different propositional languages $L_{\mathbf{Side}}$, $L_{\mathbf{Front}}$ and $L_{\mathbf{Top}}$ built up from the sets $AP_{\mathbf{Side}} = \{l, r\}$, $AP_{\mathbf{Front}} = \{l, c, r\}$, and $AP_{\mathbf{Top}} = \{a1, a2, a3, b1, b2, b3\}$ of propositional constants (where l means that the observer sees a ball in the *left* sector, c means that the observer sees a ball in the *central* sector, and so on) To account for the specific domain of ϵ, and its shift in the approximation level described above, the language L_ϵ contains a set $\{Side, Front, Top\}$ of constant symbols for each one of the contexts above, a set of constant symbol "ϕ" for each formula ϕ that can be expressed in the languages $L_{\mathbf{Side}}$ or $L_{\mathbf{Front}}$ or $L_{\mathbf{Top}}$, and a binary predicate $ist(c, "\phi")$, whose intuitive meaning is that formula $\phi \in L_c$ is true in context c (see McCarthy (1993)). In order to solve the puzzle ϵ needs to relate information contained in different contexts associated with different levels of approximations. In particular Benerecetti et al. (2000) needs to formalize the relation denoted as arrows connecting contexts in figure 6. This is done by imposing a compatibility relation between the models of each observers' context c and models of the context ϵ. To state the correspondence between a formula ϕ in each observers' context c and the formula $ist(c, "\phi")$ (denoting the same fact at a different degree of approximation) in context ϵ, if a formula of the form $ist(c, "\phi")$ is a theorem in ϵ, then the formula ϕ must be a theorem in c, and vice-versa. The different perspectival representations are formalized in Benerecetti et al. (2000) by the different (initial) axioms satisfied in each observer's context and the relations between them are explicitly stated as axioms in context ϵ.

Belief contexts. LMS and MCS have been applied to formalize different aspects of intentional contexts, and in particular belief contexts (see e.g.,Giunchiglia et al. (1993), Cimatti and Serafini (1995)). An example is a puzzle described in Benerecetti et al. (1998), where LMS and MCS are used to solve the problem of the opaque and transparent reading of belief reports.

A computer program ϵ knows that Mr. A believes that the president of the local football team is Mr. M and that Mr. B believes that the president is Mr. C. The computer program knows also that Mr. B knows that A believes that the president of the local football team is Mr. M. Actually, Mr. B is right, and the computer program knows that. Now, B tells ϵ: "Mr. A believes that the president of the local football team is a corruptor". How will ϵ interpret the sentence?

The program is a little puzzled, since the question has two possible answers: (1) Mr. A's belief is referred to Mr. M (since Mr. A is the subject of the belief). This is an instance of opaque reading. (2) Mr. A's belief is referred to Mr. C (since it is Mr. B who is speaking). This is an instance of transparent reading.

We are not interested here in the solution of the puzzle (the interested reader may refer to Benerecetti et al. (1998)), but in analyzing the representations of the different contexts involved in the formalization.

The formalization is based on the notion of *belief context*. A belief context is a representation of a collection of beliefs that a reasoner (in this example, the program) ascribes to an agent (including itself) from a given perspective. Possible perspectives are: the beliefs that the program ascribes to itself (e.g., that Mr. B believes that Mr. A believes that the president of the local football team is a corruptor); the beliefs that the program ascribes to Mr. B (e.g., that Mr. A believes that the president of the local football team is a corruptor); the beliefs that the program ascribes to Mr. B about Mr. A (e.g., that the president of the local football team is a corruptor). The belief contexts that the program can build in this example can be organized in a structure like that presented in figure 7.
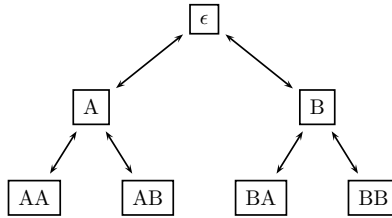
FIGURE 7 The structure of belief contexts

ϵ represents the context containing the beliefs that the program ascribes to itself, **A** is the context containing the beliefs that the program ascribes to Mr. A, **B** is the context containing the beliefs that the program ascribes to Mr. B, **BA** is the context containing the beliefs that the program ascribes to Mr. A from Mr. B's perspective, and so on.

The formalization of the different contexts in figure 7 may vary along the three dimensions of contextual dependence. Focusing on partiality, the different contexts are related to different sets of beliefs. For instance, **A** is the context containing the beliefs that the program ascribes to Mr. A, whereas **B** is the context containing the beliefs that the program ascribes to Mr. B. Focusing on approximation, we notice that each

context in the hierarchy must be able to to talk about beliefs contained in each one of the contexts above. In order to do this it needs to make explicit some information that was implicit in the observers' contexts. In particular, it needs to make explicit what beliefs are contained in what context. The relations involving different contexts associated with different degrees of approximations are the one denoted as arrows in figure 6 and are similar to the ones described in the MB example. Focusing on perspective, each of the belief contexts expresses knowledge about the world which depends on the cognitive perspective of the agents, from the point of view of the computer program. For instance, Mr. B will refer to Mr. C as "the president of the local football team", whereas Mr. A will refer to Mr. C as Mr. M.

Integration of different information sources. LMS and MCS have been applied to formalize the integration of information coming from different information sources. Ghidini and Serafini (1998a,b) contain the formal definitions and motivating examples. Let us focus on a simple example.

A mediator m of an electronic market place collects information about fruit prices from 1, 2, and 3 and integrates it in a unique homogeneous database. Customers that need information about fruit prices may therefore submit a single query to the mediator instead of contacting the sellers.

The formalization of the exchange of information in this example based on the four contexts and the information flows depicted in figure 8. Circles represent contexts associated to the different databases and arrows represent information flow between contexts (databases).
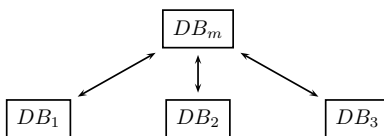
FIGURE 8 Contexts in the mediator example.

The representations of the different contexts in figure 8 may have different degrees of partiality, as each database is associated to a specific domain. For instance, the sellers might provide different subsets of fruits and therefore the domains of their databases are different. Focusing on approximation, the domain of fruits can be represented at different level of details by different sellers. E.g., database 1 may contain prices for red apples and yellow apples, while database 2 and 3 abstracts away the dependence on the color and do not make this distinction. Focusing on

perspective, prices of the different sellers might be not homogeneous, depending on their particular viewpoint. E.g., prices of database 1 don't include taxes, while prices of database 2, 3 and the mediator do.

Formally, the specific domains of the different databases are described by using different first order languages. Each database is associated with a different interpretation domain. The compatibility relation between the different levels of approximations in the fruit domains is formalized by using *domain relations*, i.e. relations between the interpretation domains of the different databases. A domain relation may, for instance, relate a "more abstract" object (e.g. apple) in the domain of a database to a set of "less abstract" objects (e.g. red−apple, green−apple) in the domain of another database. Compatibility relations between the different perspectival views contained in the databases are formalized by using *view constraints*, i.e. relations between formulae contained in different languages (databases). For instance every time the models of database 1 satisfy the formula $hasprice(x, y)$ (meaning that item x has price y, then the models of the mediator database must satisfy the formula $\exists y' hasprice(x, y') \wedge y' = y + (0.07 * y)$ (meaning that the same item x has price y' which is obtained adding the amount of taxes to y.

1.6   Conclusions

In the chapter, we have not presented a new theory about partiality, approximation, or perspective. Instead, we have shown that the work on contextual reasoning in AI (and not only in AI) can be re-read as an attempt of providing a logic of the mechanisms that govern the relationship between partial, approximate, and perspectival representations of the world.

In this sense, the work described here is only a preliminary step. Indeed, it opens a whole field of research, both philosophical and logical. Our next step will be a formal study of a logic of partiality, approximation, and perspective in the framework of LMS and MCS. In particular, we are interested in finding the compatibility relations (and the relative bridge rules) involved in the corresponding reasoning mechanisms. This, we hope, will be part of a new approach to a theory of knowledge representation, in which context will play a crucial role.

References

Attardi, G. and M. Simi. 1995. A formalisation of viewpoints. *Fundamenta Informaticae* 23(2–4):149–174.

Benerecetti, M., P. Bouquet, and C. Ghidini. 1998. Formalizing Belief Reports – The Approach and a Case Study. In F. Giunchiglia, ed., *Proceed-*

ings AIMSA'98, 8th International Conference on Artificial Intelligence, Methodology, Systems, and Applications, vol. 1480 of LNAI, pages 62–75. Sozopol, Bulgaria: Springer Verlag.

Benerecetti, M., P. Bouquet, and C. Ghidini. 2000. Contextual Reasoning Distilled. Journal of Theoretical and Experimental Artificial Intelligence 12(3):279–305.

Bouquet, P. and F. Giunchiglia. 1995. Reasoning about Theory Adequacy. A New Solution to the Qualification Problem. Fundamenta Informaticae 23(2–4):247–262. Also IRST-Technical Report 9406-13, IRST, Trento, Italy.

Buvac, Sasa. and Ian A. Mason. 1993. Propositional logic of context. Proc. of the 11th National Conference on Artificial Intelligence .

Cimatti, A. and L. Serafini. 1995. Multi-Agent Reasoning with Belief Contexts: the Approach and a Case Study. In M. Wooldridge and N. R. Jennings, eds., Intelligent Agents: Proceedings of 1994 Workshop on Agent Theories, Architectures, and Languages, no. 890 in Lecture Notes in Computer Science, pages 71–85. Springer Verlag. Also IRST-Technical Report 9312-01, IRST, Trento, Italy.

Dinsmore, J. 1991. Partitioned Representations. Kluwer Academic Publishers.

Fauconnier, G. 1985. Mental Spaces: aspects of meaning construction in natural language. MIT Press.

Ghidini, C. and F. Giunchiglia. 2001. Local models semantics, or contextual reasoning = locality + compatibility. Artificial Intelligence 127(2):221–259.

Ghidini, C. and L. Serafini. 1998a. Distributed First Order Logics. In D. Gabbay and M. de Rijke, eds., Frontiers Of Combining Systems 2, Studies in Logic and Computation, pages 121–140. Research Studies Press. ISBN 0-86380-252-4.

Ghidini, C. and L. Serafini. 1998b. Information Integration for Electronic Commerce. In C. Sierra and P. Noriega, eds., Agent Mediated Electronic Commerce. First International Workshop on Agent Mediated Electronic Trading, AMET-98, vol. 1571 of LNAI, pages 189–206. Minneapolis, USA: Springer Verlag. Also accepted for presentation at UKMAS'98. Manchester, U.K. December 14–15, 1998.

Giunchiglia, F. 1993. Contextual reasoning. Epistemologia, special issue on I Linguaggi e le Macchine XVI:345–364. Short version in Proceedings IJCAI'93 Workshop on Using Knowledge in its Context, Chambery, France, 1993, pp. 39–49. Also IRST-Technical Report 9211-20, IRST, Trento, Italy.

Giunchiglia, F. and P. Bouquet. 1997. Introduction to contextual reasoning. An Artificial Intelligence perspective. In B. Kokinov, ed., Perspectives on Cognitive Science, vol. 3, pages 138–159. Sofia: NBU Press. Lecture Notes of a course on "Contextual Reasoning" of the European Summer School on Cognitive Science, Sofia, 1996.

Giunchiglia, F. and L. Serafini. 1994. Multilanguage hierarchical logics, or: how we can do without modal logics. *Artificial Intelligence* 65(1):29–70. Also IRST-Technical Report 9110-07, IRST, Trento, Italy.

Giunchiglia, F., L. Serafini, E. Giunchiglia, and M. Frixione. 1993. Non-Omniscient Belief as Context-Based Reasoning. In *Proc. of the 13th International Joint Conference on Artificial Intelligence*, pages 548–554. Chambery, France. Also IRST-Technical Report 9206-03, IRST, Trento, Italy.

Guha, R.V. 1991. Contexts: a Formalization and some Applications. Tech. Rep. ACT-CYC-423-91, MCC, Austin, Texas.

Laird, J. E., A. Newell, and P.S. Rosenbloom. 1987. Soar: An architecture for general intelligence. *Artificial Intelligence* 33(3):1–4 64.

Lewis, D. 1980. Index, Context, and Content. In S. Kranger and S. Ohman, eds., *Philosophy and Grammar*, pages 79–100. D. Reidel Publishing Company.

McCarthy, J. 1991. Overcoming an Unexpected Obstacle. Unpublished.

McCarthy, J. 1993. Notes on Formalizing Context. In *Proc. of the 13th International Joint Conference on Artificial Intelligence*, pages 555–560. Chambery, France.

Sperber, Dan and Deirdre Wilson. 1986. *Relevance. Communication and Cognition*. Basil Blackwell.

Weyhrauch, R.W. 1980. Prolegomena to a Theory of Mechanized Formal Reasoning. *Artificial Intelligence* 13(1):133–176.

2

What is Local Models Semantics?

Chiara Ghidini, Fausto Giunchiglia

In recent papers a new semantics, called *Local Models Semantics*, was presented and used to provide a foundation to reasoning with contexts. Local Models Semantics captures and makes precise the two main principles underlying contextual reasoning: the, so-called, Principle of Locality and Principle of Compatibility. In this chapter we aim at explaining the main intuitions underlying Local Models Semantics, its fundamental logical properties, and its relation with contextual reasoning. The emphasis is on motivations and intuitions, rather than on technicalities.

2.1 Introduction

In recent papers a new semantics, called *Local Models Semantics*, was presented and used to provide a foundation to reasoning with contexts. An exhaustive presentation of the notion of context is out of the scope of this chapter.[1] The notion of context we consider here is based on two significative (informal) definitions independently proposed by Fausto Giunchiglia (1993) and John McCarthy (1993) in the late 80's, when context was introduced as an important means for formalising certain forms of reasoning.

According to Giunchiglia (1993), contexts are a tool for formalising the locality of reasoning:

> Our intuition is that reasoning is usually performed on a subset of
> the global knowledge base. The notion of context is used as a means

---

[1] The interested reader may refer to Akman and Surav (1996), Ghidini and Giunchiglia (2001), Giunchiglia (1993) for an accurate discussion on this topic.

> of formalising this idea of localisation. Roughly speaking, we take a context to be the set of facts used locally to prove a given goal plus the inference routines used to reason about them (which in general are different for different sets of facts) Giunchiglia (1993).

In McCarthy (1993), contexts are introduced as a means for solving the problem of generality:

> When we take the logic approach to AI, lack of generality shows up in that the axioms we devise to express common sense knowledge are too restricted in their applicability for a general common sense database [...] Whenever we write an axiom, a critic can say that the axiom is true only in a certain context. With a little ingenuity the critic can usually devise a more general context in which the precise form of the axiom doesn't hold. McCarthy (1987)

Coherently with these two proposals, contexts have been used in various applications and in different domains. Contexts are used to deal with issues concerning the integration of heterogeneous knowledge and data bases. See for instance Farquhar et al. (1995), Mylopoulos and Motschnig-Pitrik (1995), Ghidini and Serafini (1998b), Theodorakis et al. (1998). The largest common-sense knowledge-base, CYC Lenat (1995), contains an explicit notion of context Guha (1991). Several references can be found in the literature about the use of contexts in the formalisation of reasoning about beliefs, meta reasoning, and propositional attitudes. See for instance Giunchiglia et al. (1993), Giunchiglia and Giunchiglia (1996), Benerecetti et al. (1998a), Fisher and Ghidini (1999), Ghidini (1999), Giunchiglia and Serafini (1994). In Attardi and Simi (1995) contexts are introduced in the formalisation of reasoning with viewpoints. Bouquet and Giunchiglia (1995) addresses the problem of formalising context-based common-sense reasoning. Finally, Benerecetti et al. (1998b), Cimatti and Serafini (1995), Parsons et al. (1998), Noriega and Sierra (1996), Ghidini and Serafini (1998a) introduce contexts to model different aspects of agents and multi-agent systems.

In spite of the variety of different approaches, formalisations, and domains of application, in Ghidini and Giunchiglia (2001) the authors claim that there are two main intuitions underlying the use of context, and state them as the following two principles:

Principle 1 (of Locality): reasoning uses only part of what is potentially available (e.g., what is known, the available inference procedures). The part being used while reasoning is what we call *context* (of reasoning);

Principle 2 (of Compatibility): there is *compatibility* among the

reasoning performed in different contexts.

Local Models Semantics provides a formal framework where the two principles of Locality and Compatibility are captured and made precise. The goal of this chapter is to explain the main intuitions underlying Local Models Semantics, its fundamental logical properties, and its relations with contextual reasoning. The emphasis is on motivations and intuitions, rather than on technicalities. The reader interested in a more technical presentation and a detailed comparison with other logical frameworks may refer to Ghidini and Giunchiglia (2001)

The chapter is organised as follows. The core definitions are given in Sections 2.3 and Section 2.4. In Section 2.5 we comment on the properties of Local Models Semantics. In particular we investigate how the notion of context is formally defined within Local Models Semantics, and how Local Models Semantics captures the principles of Locality and Compatibility introduced above. In Section 2.6 we comment on how Local Models Semantics is able to deal with situations where we may or may not have a complete description of the world. To make the presentation clearer, in Section 2.2 we introduce a simple example of reasoning with viewpoints, called the *magic box* example, which will be used throughout the chapter. This example is a variation of the one originally proposed in Ghidini and Giunchiglia (2001).

2.2   The magic box example

Suppose there are two observers, Mr. Blue and Mr. Pink, each having a partial view of a box as shown in Figure 1. The box is composed of six

FIGURE 1   The magic box.
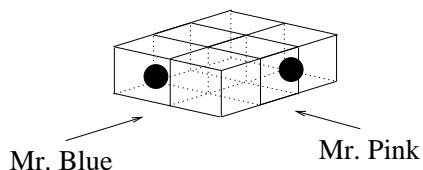
sectors, each sector possibly containing a ball. There must be exactly two balls in the box and there cannot be balls hidden from the view of an observer. The box is "magic" and observers cannot distinguish the depth inside it. Figure 2 shows the views of Mr. Blue and Mr. Pink corresponding to the scenario depicted in Figure 1.

In this example we focus on the two contexts describing the view-
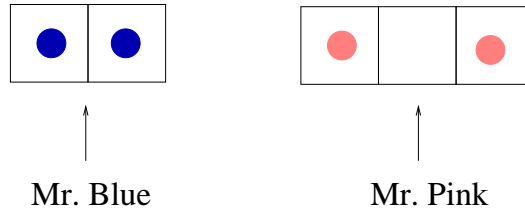
Mr. Blue   Mr. Pink

FIGURE 2 The contexts of Mr. Blue and Mr. Pink.

points of the two observers and the consequences that they are able
to draw from it. The content of the two contexts corresponding to the
scenario depicted in Figure 1 is graphically represented in Figure 2.

It is easy to see that the notions of locality and compatibility play a
central role in this example. First locality. Both Mr. Blue and Mr. Pink
have the notions of a ball being on the right or on the left. However
we may have a ball which is on the right for Mr. Blue and not on the
right for Mr. Pink. Furthermore Mr. Pink has the notion of "a ball
being in the center of the box" which is meaningless for Mr. Blue. We
also assume that the box is made of different coloured glass. Different
observers, looking at the box from different sides, see the balls as if in
different colours. In our example Mr. Blue sees (has the notion of) a
ball being blue, while Mr. Pink sees (has the notion of) a ball being
pink.

Focusing on compatibility, the contents of the contexts of Mr. Blue
and Mr. Pink are obviously related. The relation is a consequence of
the fact that Mr. Blue and Mr. Pink see the same box. Given the fact
that there must be exactly two balls in the box, it is easy to see that
if Mr. Blue sees only one blue ball in the box, then Mr. Pink must
see two pink balls in the box. Therefore we can describe this situation
by listing all the possible compatible pairs (as they are represented in
Figure 3), or we can describe it more synthetically using descriptions
like: "if Mr. Blue sees a single blue ball then Mr. Pink sees two pink
balls" and "if Mr. Pink sees a single pink ball then Mr. Pink sees two
blue balls".

2.3   Local models and model

We begin here the presentation of Local Models Semantics by defining
the notions of local model and model.

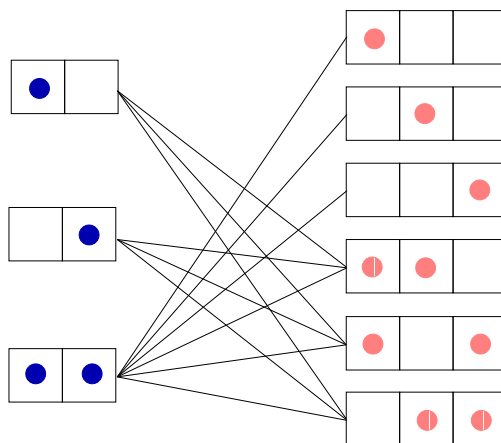Mr. Blue's contexts   Mr. Pink's contexts

FIGURE 3 Compatible contexts of Mr. Blue and Mr. Pink.

2.3.1 The formal definitions

Let $\{L_i\}_{i \in I}$ be a family of languages defined over a set of indexes I (in the following we drop the index $i \in I$). Intuitively, each L_i is the formal language used to describe what is true in a context. For the purpose of our work we suppose that I is at most countable and that $\{L_i\}$ is a class of propositional languages. The first step towards the definition of a model for $\{L_i\}$ is to consider the class of models for each language L_i in $\{L_i\}$. This will ensure that each language L_i is interpreted in its own, possibly different, structure. Formally, we denote with $\overline{M}_i$ the class of all the models of L_i. We call $m \in \overline{M}_i$ a *local model* (of L_i).

Then, we have to pair local models into a single structure. This is done by introducing the notions of compatibility sequence and compatibility relation. Formally, a *compatibility sequence* $\mathbf{c}$ (for $\{L_i\}$) is a sequence

$$\mathbf{c} = \langle \mathbf{c}_0, \mathbf{c}_1, \dots, \mathbf{c}_i, \dots \rangle$$

where, for each $i \in I$, $\mathbf{c}_i$ is a subset of $\overline{M}_i$. We call $\mathbf{c}_i$ the i-th element of $\mathbf{c}$. If $I = \{1, 2\}$ is composed of two indexes, a compatibility sequence $\mathbf{c}$ is of the form $\mathbf{c} = \langle \mathbf{c}_1, \mathbf{c}_2 \rangle$ and is called a *compatibility pair*.

A *compatibility relation* $\mathbf{C}$ (for $\{L_i\}$) is a set $\mathbf{C} = \{\mathbf{c}\}$ of compatibility sequences $\mathbf{c}$.[2]

---

[2]Formally, let $\prod_{i \in I} 2^{\overline{M}_i}$ be the Cartesian product of the collection $\{2^{\overline{M}_i} : i \in I\}$.

We define a model as a compatibility relation which contains at least one sequence and does not contain the sequence of empty sets. Formally, a *model* (for $\{L_i\}$) is a compatibility relation $\mathbf{C}$ such that:

1. $\mathbf{C} \neq \emptyset$;
2. $\langle \emptyset, \emptyset, \ldots, \emptyset, \ldots \rangle \notin \mathbf{C}$.

In the following we write $\mathbf{C}$ to mean either a compatibility relation or a model, the context always makes clear what we mean.

In a nutshell, we can split the construction we perform into three steps. First, we start with some language, say $L_1, L_2,$ and L_3 (see Figure 4). Then, we associate each L_i with a set $M_i \subseteq \overline{M}_i$ of local models. Usually $M_i \subset \overline{M}_i$ (see Figure 5). Finally, we pair local models inside compatibility sequences. The resulting compatibility relation is our model (see Figure 6).[3] Local models describe what is locally true. Compatibility sequences put together local models which are "mutually compatible", consistently with the situation we are describing. What we obtain are models composed of sets of "mutually compatible" sequences of local models.

Given a family of languages $\{L_i\}$, different classes of models may be defined, depending on the definition of compatibility relation. Different compatibility relations model different situations. A general class of models which will be used often in the chapter is based on the notion of chain. A compatibility sequence $\mathbf{c}$ is a *chain* if all the $\mathbf{c}_i$ contain exactly one local model (formally, if $|\mathbf{c}_i| = 1$ for each $i \in I$). A model $\mathbf{C}$ is a *chain model* if all the $\mathbf{c}$ in $\mathbf{C}$ are chains.

2.3.2 A model for the magic box

Let us apply the three step construction of the model depicted in Figures 4, 5, and 6 to the magic box example.

Languages We define the propositional languages L_B and L_P used by Mr. Blue and Mr. Pink, respectively, to describe their views. Let $P_B = \{r, l\}$ and $P_P = \{r, c, l\}$ be two sets of propositional constants. Intuitively, r, c, l stand for ball on the right, in the center and on the left, respectively. L_B is formally defined as the smallest set containing P_P, the symbol for falsity $\perp$, and closed under implication; L_P is formally defined as the smallest set containing P_2, the symbol for falsity $\perp$ and closed under implication. In this chapter we use the standard abbreviations from propositional logic, such as $\neg\phi$ for $\phi \supset \perp$, $\phi \vee \psi$ for $\neg\phi \supset \psi$, $\phi \wedge \psi$ for $\neg(\neg\phi \vee \neg\psi)$, $\top$ for $\perp \supset \perp$.

The compatibility relation $\mathbf{C}$ is a relation of type $\mathbf{C} \subseteq \prod_{i \in I} 2^{\overline{M}_i}$

[3]Figures 4, 5, and 6 first appeared in Ghidini and Giunchiglia (2001).

FIGURE 4 Languages: L_1, L_2, and L_3.
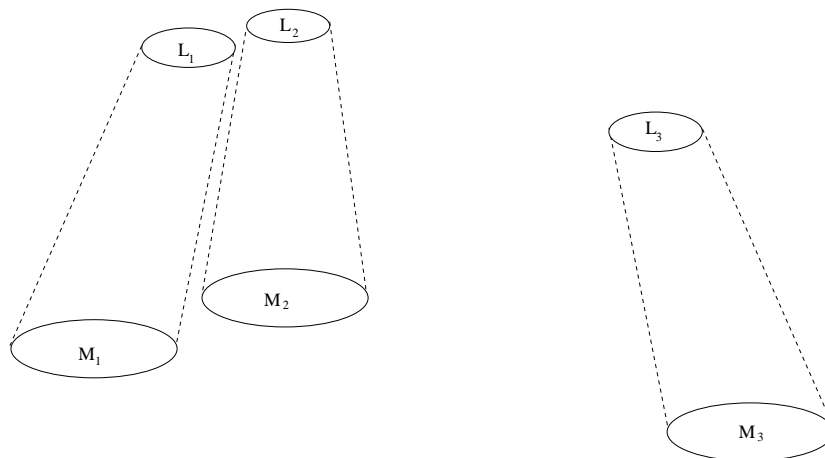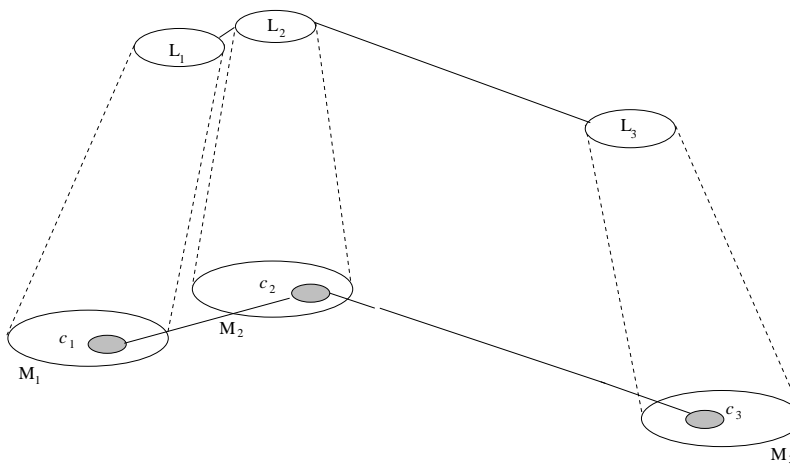

FIGURE 5 Local models for L_1, L_2, and L_3.

FIGURE 6 Model for $\{L_1, L_2, L_3\}$.

Local models We construct all the possible situations (local models) for L_B and L_P. L_B and L_P have the usual propositional semantics. Therefore the local models of L_B and L_P are univocally defined by sets of propositional formulae. In particular, the local models of L_B are univocally denoted by the following sets of formulae:

$$m_1 = \{l\} \qquad m_2 = \{r\} \qquad m_3 = \{l, r\}$$

where we write $\{l\}$ to mean the local model describing the situation with a ball on the left, $\{r\}$ to mean the local model describing the situation with a ball on the right, and $\{l, r\}$ describing the situation with a ball on the left and a ball on the right.

Analogously, the local models of L_P are univocally denoted by the following sets of formulae:

$$m_1 = \{l\} \qquad m_2 = \{c\} \qquad m_3 = \{r\}$$
$$m_4 = \{l, c\} \qquad m_5 = \{l, r\} \qquad m_6 = \{c, r\}.$$

Remember that there must be exactly two balls in the magic box. For this reason $\{l, c, r\}$ is not a local model describing a viewpoint of Mr. Pink.

Compatibility relations and model Following the definition given in Section 2.3, a generic compatibility pair for the magic box is a pair $\langle \mathbf{c}_B, \mathbf{c}_P \rangle$ where $\mathbf{c}_B$ is a set of models of the view of Mr. Blue and $\mathbf{c}_P$ is a set of models of the view of Mr. Pink. A model is a set of compatibility pairs.

In order to construct a model for the scenario described in Figure 3 (Section 2.2), we impose the following compatibility constraints:

$$\text{if Mr. Blue sees a single blue ball}$$
$$\text{then Mr. Pink sees two pink balls} \tag{2.1}$$

$$\text{if Mr. Pink sees a single pink ball}$$
$$\text{then Mr. Blue sees two blue balls} \tag{2.2}$$

$$\text{Mr. Blue and Mr. Pink are able to construct}$$
$$\text{a complete description of their view} \tag{2.3}$$

Notationally we use the following shorthand:

- *one*(l, r) for $(l \lor r) \land \neg(l \land r)$;
- *one*(l, c, r) for $(l \lor c \lor r) \land \neg(l \land r) \land \neg(l \land c) \land \neg(c \land r)$;
- *two*(l, c, r) for $((l \land r) \lor (l \land c) \lor (c \land r)) \land \neg(l \land c \land r)$.

Constraints (2.1)-(2.3) are captured, at a formal level, by the following definition. A model **C** for the magic box is a compatibility relation

such that, for all $\mathbf{c} \in \mathbf{C}$

$$\text{if } \mathbf{c}_B \text{ satisfies } one(l,r) \text{ then } \mathbf{c}_P \text{ satisfies } two(l,c,r) \qquad (2.4)$$

$$\text{if } \mathbf{c}_P \text{ satisfies } one(l,c,r) \text{ then } \mathbf{c}_B \text{ satisfies } l \wedge r \qquad (2.5)$$

$$|\mathbf{c}_B| = 1 \text{ and } |\mathbf{c}_P| = 1 \qquad (2.6)$$

Let us explore in detail the relation between the informal compatibility constraints (2.1)-(2.3) and Equations (2.4)-(2.6). Equation (2.4) models constraint (2.1). In fact, if Mr. Blue sees a ball then this ball can be on the left or on the right and the formula $one(l,r)$ describes his view. Furthermore, in this case, Mr. Pink sees two balls in two of the three possible positions, and, therefore $two(l,c,r)$ represents his view. A similar explanation can be given for Equation (2.5), which models constraint (2.2). Equation (2.6) is more interesting. It says that $\mathbf{c}_B$ and $\mathbf{c}_P$ contain a single local model, i.e., the magic box model is a chain model. This intuitively means that both Mr. Blue and Mr. Pink see the box (from their point of view) and are able to construct a complete description of it. As a consequence of Equation (2.6), a model $\mathbf{C}$ for the magic box example in Figure 3 is a set of pairs $\langle \{m_B\}, \{m_P\} \rangle$ where m_B and m_P are local models of L_B and L_P, respectively. Each pair corresponds to a possible combination of the observers' partial views. The model $\mathbf{C}$ containing all and only the compatibility pairs depicted in Figure 3 is represented in Equation (2.7). All the models satisfying Equations (2.4)-(2.6) are subsets of this model.

$$\mathbf{C} = \left\{ \begin{array}{ll} \langle\{l\},\{l,c\}\rangle, & \langle\{l\},\{l,r\}\rangle, \\ \langle\{l\},\{c,r\}\rangle, & \langle\{r\},\{l,c\}\rangle, \\ \langle\{r\},\{l,r\}\rangle, & \langle\{r\},\{c,r\}\rangle, \\ \langle\{l,r\},\{l\}\rangle, & \langle\{l,r\},\{c\}\rangle, \\ \langle\{l,r\},\{r\}\rangle, & \langle\{l,r\},\{l,c\}\rangle, \\ \langle\{l,r\},\{l,r\}\rangle, & \langle\{l,r\},\{c,r\}\rangle \end{array} \right\} \qquad (2.7)$$

As a final remark notice that linking local models inside a model may force us to eliminate some of them. Suppose that we restrict ourselves to consider local models for Mr. Blue which allow for exactly one ball. This leads to the definition of the two local models $\{l\}$ and $\{r\}$ for L_B depicted on the lefthand side in Figure 7, and of the six possible local models $\{l\}, \{c\}, \{r\}, \{l,c\}, \{l,r\}, \{c,r\}$ for L_P depicted on the righthand side in Figure 7. We know that if Mr. Blue sees a single ball, then Mr. Pink must see two balls. As a consequence, the model for the situation in which Mr. Blue sees exactly one ball does not contain any pair, and corresponding local models for Mr. Pink, which represent that

28 / Chiara Ghidini, Fausto Giunchiglia

Mr. Blue's contexts   Mr. Pink's contexts
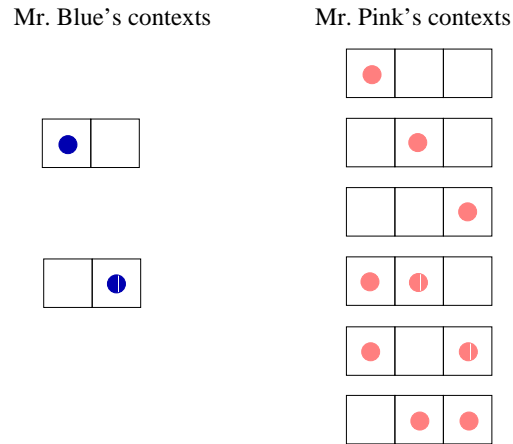
FIGURE 7 Mr. Blue sees exactly one ball: the local models.

Mr. Pink sees a single ball. The resulting model is indeed the following:

$$\left\{ \begin{array}{ll} \langle\{l\},\{l,c\}\rangle, & \langle\{l\},\{l,r\}\rangle, \\ \langle\{l\},\{c,r\}\rangle, & \langle\{r\},\{l,c\}\rangle, \\ \langle\{r\},\{l,r\}\rangle, & \langle\{r\},\{c,r\}\rangle \end{array} \right\}$$

and is graphically represented in Figure 8.

Mr. Blue's contexts   Mr. Pink's contexts
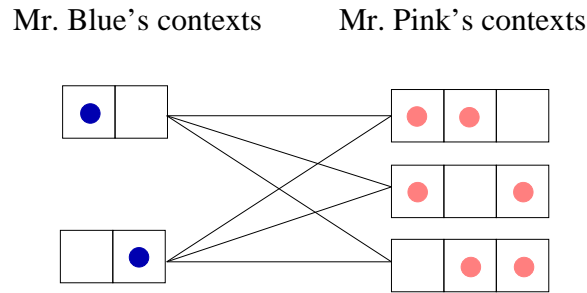
FIGURE 8 Mr. Blue sees exactly one ball: the model.

2.4 Satisfiability and logical consequence

The definition of satisfiability of a formula of a language L_i in the model $\mathbf{C}$, is based on the satisfiability of the same formula in the local models

of L_i. Formally, let $\models_{cl}$ be the satisfiability relation between local models and formulae of L_i. We call $\models_{cl}$ *local satisfiability*. Notationally, let us write $i\!:\!\phi$ to mean ϕ, where ϕ is a formula of L_i. We say that ϕ is an L_i-formula, and that $i\!:\!\phi$ is a formula or, also, a labelled L_i-formula. This notation and terminology allows us to keep track of the context we are talking about.

Let $\mathbf{C} = \{\mathbf{c}\}$ with $\mathbf{c} = \langle \mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_i, \ldots \rangle$ be a model and $i : \phi$ a formula. $\mathbf{C}$ *satisfies* $i\!:\!\phi$, in symbols $\mathbf{C} \models i\!:\!\phi$, if for all $\mathbf{c} \in \mathbf{C}$

$$\mathbf{c}_i \models \phi$$

where $\mathbf{c}_i \models \phi$ if, for all $m \in \mathbf{c}_i$, $m \models_{cl} \phi$.

The intuition underlying the motion of satisfiability is that an L_i-formula is satisfied by a model $\mathbf{C}$ if all the local models in each $\mathbf{c}_i$ satisfy it.

Consider, for instance, the simple model

$$\mathbf{C}' = \{ \ \langle \{l\}, \{c, r\} \rangle, \quad \langle \{l, r\}, \{l, c\} \rangle \ \} \tag{2.8}$$

containing only the two compatibility pairs depicted in Figure 9. According to the definition of satisfiability $\mathbf{C}'$ satisfies the formula $B\!:\!l$, meaning that Mr. Blue sees a ball in the left position. This is because the two local models $\{l\}$ and $\{l, r\}$ for L_B contained in $\mathbf{C}'$ both satisfy the formula l. On the contrary, $\mathbf{C}'$ does not satisfy $B\!:\!r$, meaning that Mr. Blue sees a ball in the right position. This is because there is a local model for Mr. Blue, namely $\{l\}$, which does not satisfy the formula r.
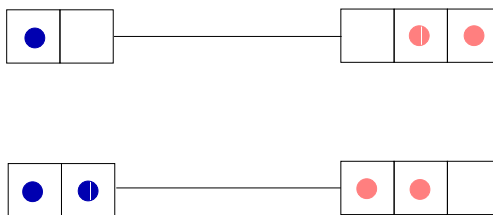
FIGURE 9 Mr. Blue sees a ball on the left.

The notions of satisfiability of a set of formulae and of validity are the obvious ones. A model $\mathbf{C}$ satisfies a set of formulae Γ, in symbols $\mathbf{C} \models \Gamma$, if $\mathbf{C}$ satisfies every formula $i\!:\!\phi$ in Γ. A formula $i\!:\!\phi$ is *valid*, in symbols $\models i\!:\!\phi$, if all models satisfy $i\!:\!\phi$.

An interesting notion is the one of *logical consequence* which must take into account the fact that assumptions and conclusion may belong to distinct languages. Given a set of labelled formulae Γ, Γ_j denotes

the set of formulae $\{\gamma \mid j\!:\!\gamma \in \Gamma\}$. A formula $i\!:\!\phi$ is a logical consequence of a set of formulae Γ w.r.t. a model $\mathbf{C}$, in symbols $\Gamma \models_{\mathbf{C}} i\!:\!\phi$, if every sequence $\mathbf{c} \in \mathbf{C}$ satisfies:

$$\forall j \in I, j \neq i, \ \mathbf{c}_j \models \Gamma_j \Longrightarrow (\forall m \in \mathbf{c}_i, \ m \models_{cl} \Gamma_i \Longrightarrow m \models_{cl} \phi) \quad (2.9)$$

Equation (2.9) looks slightly complicated. Let us illustrate it with the help of an example. Consider the model of the magic box informally depicted in Figure 3 and formally represented by Equation (2.7). We want to verify that in this model

(10)   if Mr. Blue sees a ball on the left and no ball on the right, and Mr. Pink doesn't see any ball in the center, then Mr. Pink sees a ball on the left and a ball on the right.

Formally, the sentence (10) can be rewritten as

$$B\!:\!l \wedge \neg r, P\!:\!\neg c \models_{\mathbf{C}} P\!:\!l \wedge r$$

The set of assumption Γ contains the facts that "Mr. Blue sees a ball on the left and no ball on the right" and "Mr. Pink doesn't see any ball in the center". Formally, $\Gamma = \{B : l \wedge \neg r, P : \neg c\}$. The first step is to isolate the set of assumptions which are made in a context different from the context of Mr. Pink. That is $B : l \wedge \neg r$. Then we restrict ourselves to considering all the compatibility pairs whose local models satisfy the formula $B : l \wedge \neg r$, and throw away all the others. The remaining compatibility pairs are

$$\langle \{l\}, \{l, c\} \rangle$$
$$\langle \{l\}, \{l, r\} \rangle$$
$$\langle \{l\}, \{c, r\} \rangle$$

and are depicted in Figure 10. Consider now the local models of Mr. Pink in the remaining sequences. We have to identify all the local models of Mr. Pink in the remaining pairs such that there is no ball in the center. Formally, we have to identify all the local models of Mr. Pink satisfying $P : \neg c$. The only local model satisfying that Mr. Pink doesn't see any ball in the center is

$$\{l, r\}$$

and is depicted in Figure 11. The last step is to check whether the remaining local models of Mr. Pink represent the fact that Mr. Pink sees a ball on the left and a ball on the right. It is easy to see that the only remaining local model in Figure 11 satisfies this property. Therefore the model depicted in Figure 3 and formally defined in Equation (2.7) satisfies the sentence (10).

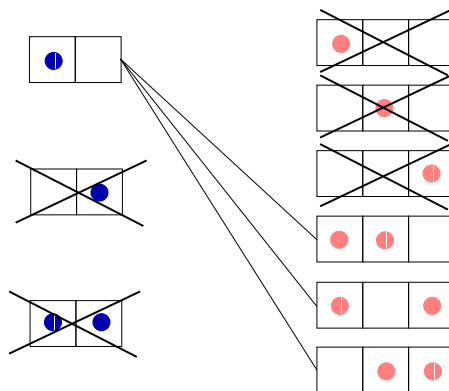Mr. Blue's contexts   Mr. Pink's contexts

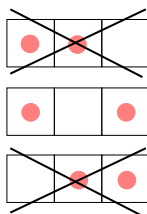FIGURE 10 Selecting compatibility sequences.

FIGURE 11 Selecting local models.

The extension of the notion of logical consequence to a class of models is the usual one. A formula $i\!:\!\phi$ is a logical consequence of a set of formulae Γ w.r.t. a class of models M, in symbols $\Gamma \models_\mathsf{M} i\!:\!\phi$, if $i\!:\!\phi$ is a logical consequence of Γ w.r.t. all the models in M. Finally, a formula $i\!:\!\phi$ is a *logical consequence* of Γ, in symbols $\Gamma \models i\!:\!\phi$, if $i\!:\!\phi$ is a logical consequence of Γ w.r.t. all models $\mathbf{C}$.

2.5 Contexts, locality and compatibility

Having formally defined the logical framework, the question now is: where are contexts in this picture? How does Local Models Semantics relate to contextual reasoning? We already suggested part of the answer to this question by illustrating the main notions of model and satisfiability using the magic box example. In this section we answer these questions in more detail by illustrating how the notion of context can be formally introduced in the framework of Local Models Semantics. We then examine how Local Models Semantics formally captures the notions of locality and compatibility.

Given a model $\mathbf{C} = \{\langle \mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_i, \ldots \rangle\}$ we formally define a *context* to be any $\mathbf{c}_i$, namely the set of local models $m \in \overline{M}_i$ allowed by $\mathbf{C}$ within any particular compatibility sequence. For instance, the contexts for Mr. Blue allowed by the model $\mathbf{C}'$ defined in Equation (2.8) are $\{l\}$ and $\{l, r\}$.

The intuition underlying the definition of context is that a context consists of that set of models which capture exactly those facts which are locally true, given also the constraints posed by the local models of other contexts in the same compatibility sequence. This notion of context is the semantic formalisation of the notion of context intuitively introduced in Principle 1 in Section 9.1.

An interesting property of this definition is that contexts are formalised as partial objects, as explicitly required in, e.g., Giunchiglia (1993), McCarthy (1987). This is due to the fact that context is defined as a set of models instead of a single model. In order to illustrate the advantage of having contexts as partial objects consider the slightly modified magic box scenario depicted in Figure 12, where Mr. Pink is able to see only one box sector and knows that there are two sectors behind the wall. In this scenario Mr. Pink is able to distinguish only two situations: there is a ball on the left, and there is no ball on the left. The fact that Mr. Pink is uncommitted to whether there is a ball in a sector behind the wall is formalised by having the sentences "there is a ball on the right" and "there is a ball in the center" true in some local models representing the view of Mr. Pink and false in others. In
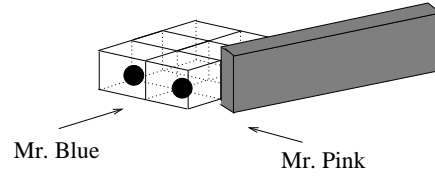
FIGURE 12 A partially hidden magic box.

the resulting context, describing the viewpoint of Mr. Pink, "there is a ball on the right" and "there is a ball in the center" will be neither true or false because there will be models in $\mathbf{c}_P$ where these sentences are false and others where the same sentences are true. Formally, the model for the scenario depicted in Figure 12 is defined as follows

$$\mathbf{C}^* = \left\{ \begin{array}{l} \langle\{l\}, \{\{c\}, \{r\}, \{c, r\}\}\rangle, \\ \langle\{r\}, \{\{c\}, \{r\}, \{c, r\}\}\rangle, \\ \langle\{l, r\}, \{\{c\}, \{r\}, \{c, r\}\}\rangle, \\ \langle\{l\}, \{\{l\}, \{l, c\}, \{l, r\}\}\rangle, \\ \langle\{r\}, \{\{l\}, \{l, c\}, \{l, r\}\}\rangle, \\ \langle\{l, r\}, \{\{l\}, \{l, c\}, \{l, r\}\}\rangle \end{array} \right\} \tag{2.11}$$

and is graphically represented in Figure 13. It is easy to see that the two contexts for Mr. Pink allowed by the model $\mathbf{C}^*$ are $\{\{c\}, \{r\}, \{c, r\}\}$ and $\{\{l\}, \{l, c\}, \{l, r\}\}$. In these contexts the formulae r and c are neither true or false. Consider, for instance, the context $\{\{c\}, \{r\}, \{c, r\}\}$ and the formula r. r is neither true or false in $\{\{c\}, \{r\}, \{c, r\}\}$ because there is a local model $\{c\}$ where r is false and another local model $\{r\}$ where r is true.

Given the above notion of context, we can now better illustrate the intuitions underlying the notions of compatibility sequence, compatibility relation, and model. A context is a partial description of the world. A compatibility sequence contains as many contexts as needed, one for each partial description of the world. Thus, in the magic box scenario we have compatibility sequences of length two, containing a context for the view of Mr. Blue and a context for the view of Mr. Pink. In the more general scenario involving n observers, we have to consider sequences of length n.

An interesting set of compatibility sequences is the one composed by chains introduced at the end of Section 2.3. Remember that a chain is a compatibility sequence in which all the contexts are singleton sets. In

Mr. Blue's contexts   Mr. Pink's contexts
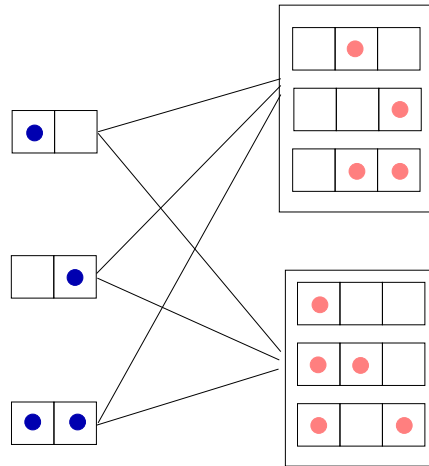
FIGURE 13   Model for the scenario of Figure 12.

this case, all the contexts are complete objects in the sense that each context, being a single model, assigns a truth value to all sentences in its language. In other words, a context which is a singleton set models the situation where a partial description of the world assigns a truth value to all the propositions it is able to express in its local (and limited) language. This is the case in Figures 1, 2, and 3. Here, Mr. Blue and Mr. Pink have partial views of the world. However, within their partial views, they are able to "see everything". On the contrary, this is not the case in Figures 12 and 13. Here, Mr. Blue is still able to "see everything" within its partial views, while Mr. Pink is not.

Local Models Semantics completely embraces the principle of Locality. We can easily say that everything is local. First of all, the languages are local to the contexts. Second, the languages are interpreted in local structures (or local models). This reflects the fact that contexts can have their own, generally different, domains of interpretation, sets of relations, and sets of functions. Third, the notion of satisfiability is local: the satisfiability of a (labelled) formula is given in terms of the local satisfiability of the formula with respect to its context.

Because of compatibility sequences, contexts mutually influence themselves. Compatibility has the structural effect of changing the set of local models defining each context. It forces local models to agree up to a certain extent. A typical example is the one depicted in

Figure 8, where the fact that Mr. Blue sees exactly a ball forces us to throw away all the pairs, and corresponding local models for Mr. Pink, which allow for zero balls.

2.6 From contexts to the world

In learning about our approach to the formalisation of the magic box example, the reader might object that the most straightforward formalisation of this example would be a direct axiomatisation of the box as a two-dimensional grid. The contexts representing the views of Mr. Blue
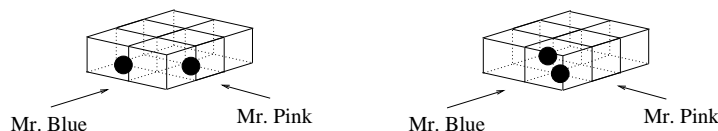
FIGURE 14 Indistinguishable situations.

and Mr. Pink could then easily be constructed by projecting the grid in two one-dimensional views. Locality and compatibility would be guaranteed by construction. However this approach is based on the hypothesis that we have a complete description of the world (the box in this case), and that we can use it to build views of the world itself. This is not always the case. Quite often we have only partial views and it is possible that we are not able to reconstruct the complete description of the world starting from the partial views, but only a partial or approximate description of it. As an example, consider the situations depicted in Figure 14. These two different situations cannot be distinguished by the two observers. That is, even assuming the existence of a third agent who knows the actual form of the box, (s)he is not able to identify which situation, among the ones depicted in Figure 14, is the current one, knowing only what Mr. Blue and Mr. Pink see. In fact, the unique pair of compatible contexts associated to the two different situations in Figure 14 is the one depicted in Figure 15.

The capability of dealing with situations where we may or may not have a complete description of the world is quite important in several application domains. Among the most important is the development and integration of data or knowledge bases. In a relational, possibly distributed, data base there is what is assumed to be a complete description of the world, and views are built by filtering out, and appropriately merging together, part of the available information. On the other hand, a federation of heterogeneous data or knowledge bases,

Mr. Blue's context   Mr. Pink's context
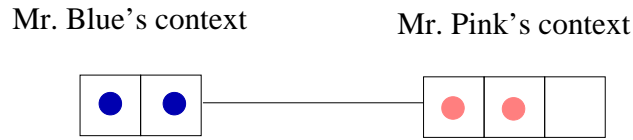
FIGURE 15 Compatible contexts in the scenario of Figure 14.

possibly developed independently, can be seen as a set of views of an ideal data base which is often impossible or very complex to reconstruct completely.

An exhaustive investigation on the relation between partial views and a complete description of the world is out of the scope of this chapter. Our aim here is to highlight the problem and suggest how Local Models Semantics is able to deal with situations where we may or may not have a complete description of the world in (simple) scenarios from the magic box example. In order to do that, consider the following scenario. The box is the same as the one depicted in Figure 1, but this time the balls have to be placed in the same column (i.e., there cannot be balls on a diagonal line). Figure 16 shows all the possible configurations allowed in this scenario from a top view of the box.
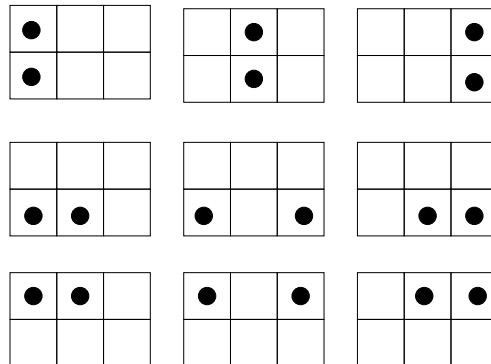
FIGURE 16 A new magic box.

It is very easy to show that in this case the observers can distinguish between all the possible situations. Figure 17 graphically describes the compatibility pairs involving the three different possible situations for Mr. Blue and the six different possible situations for Mr. Pink.

The graphical model depicted in Figure 17 doesn't look very differ-

Mr. Blue's contexts　　Mr. Pink's contexts
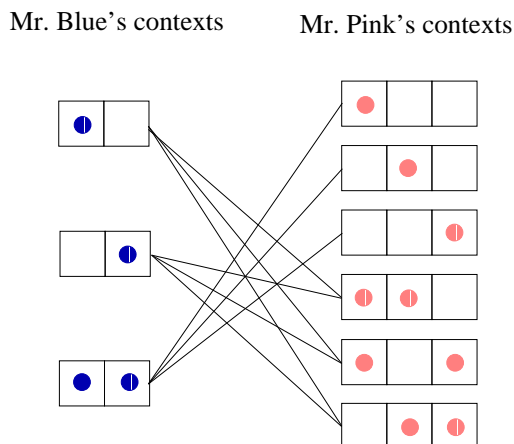
FIGURE 17 Compatible contexts in the scenario of Figure 16.

ent from the one depicted in Figure 3. So, why in this case the observers are able to distinguish between all the possible situations? Because in this case it is possible to find a precise correspondence between the compatibility pairs in Figure 17 and the complete description of the box provided by the top views in Figure 16. More formally, it is possible to find a bijective[4] function f from the set of compatibility pairs $\mathbf{C}$, graphically defined in Figure 17, to the set of models graphically defined in Figure 16. This function enables a one-to-one correspondence between every compatibility pair in Figure 17 and one of the possible descriptions of the box, in Figure 16. Figure 18 provides a graphical description of f.

Let $\mathbf{C}$ be a compatibility relation and M a set of models intuitively representing a complete description of the world. We believe that the capability of defining a bijective function f from $\mathbf{C}$ to M is a necessary condition for stating that $\mathbf{C}$ enables the reconstruction of a complete description of the world. Is this condition also a sufficient one? Due to the infinite varieties of relations existing between different views of the world we are not able to give a definite answer in this chapter. Nonetheless, one-to-one functions can provide a preliminary mechanism for controlling whether a certain model $\mathbf{C}$ provides a description

---

[4]Formally, a function f from a set A to a set B is *injective* if each element of A maps onto a different element of B. A function f from set A onto B is called *surjective* (or 'onto') if every member of B is the image of at least one member of A. A function f is *bijective* if it is both injective and surjective.

Compatibility pairs   Top views
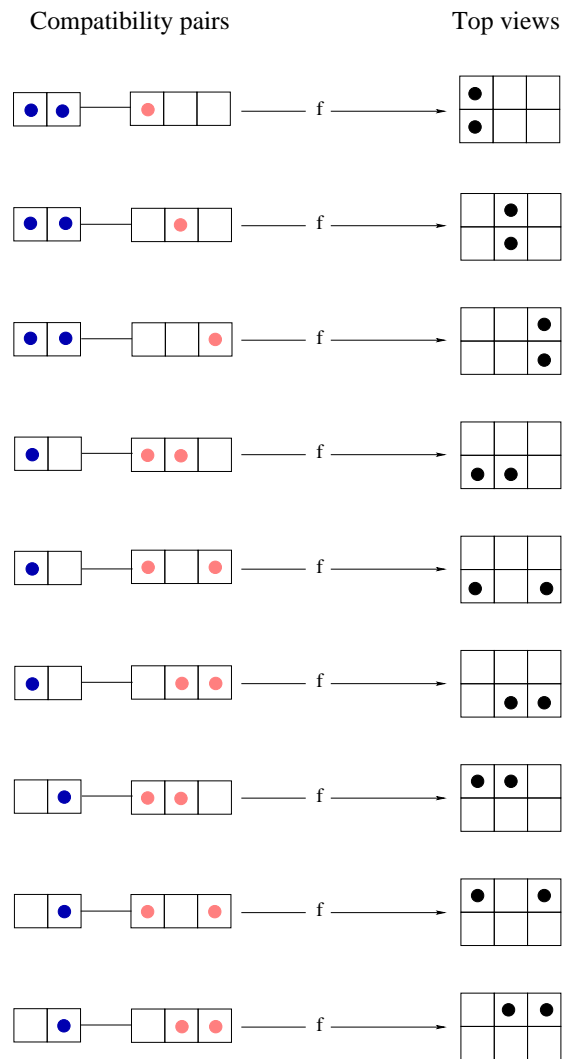
FIGURE 18 One-to-one correspondence.

of different views of the world which enables the reconstruction of a complete description of the world.

2.7 Conclusion

In this chapter we have explained a new semantics, called Local Models Semantics, which was recently proposed as a foundation to reasoning with context. Local Models Semantics formalises the two general principles underlying contextual reasoning, namely the principle of locality and the principle of compatibility. We have also shown how Local Models Semantics can be used to model a characteristic example of reasoning with viewpoints: the magic box example. Chapter 1 (Section 1.5) contains a brief description of two additional, and very different, areas where Local Models Semantics has been successfully applied: the modelling of intentional context, and belief context in particular, and the representation of semantic heterogeneity issues in information integration.

References

Akman, V. and M. Surav. 1996. Steps toward Formalizing Context. *AI Magazine* pages 55–72.

Attardi, G. and M. Simi. 1995. A formalisation of viewpoints. *Fundamenta Informaticae* 23(2–4):149–174.

Benerecetti, M., P. Bouquet, and C. Ghidini. 1998a. Formalizing Belief Reports – The Approach and a Case Study. In F. Giunchiglia, ed., *Proceedings AIMSA'98, 8th International Conference on Artificial Intelligence, Methodology, Systems, and Applications*, vol. 1480 of *LNAI*, pages 62–75. Sozopol, Bulgaria: Springer Verlag.

Benerecetti, M., F. Giunchiglia, and L. Serafini. 1998b. Model Checking Multiagent Systems. *Journal of Logic and Computation, Special Issue on Computational & Logical Aspects of Multi-Agent Systems* 8(3):401–423.

Bouquet, P. and F. Giunchiglia. 1995. Reasoning about Theory Adequacy. A New Solution to the Qualification Problem. *Fundamenta Informaticae* 23(2–4):247–262.

Cimatti, A. and L. Serafini. 1995. Multi-Agent Reasoning with Belief Contexts: the Approach and a Case Study. In M. Wooldridge and N. R. Jennings, eds., *Intelligent Agents: Proceedings of 1994 Workshop on Agent Theories, Architectures, and Languages*, no. 890 in Lecture Notes in Computer Science, pages 71–85. Springer Verlag.

Farquhar, A., A. Dappert, R. Fikes, and W. Pratt. 1995. Integrating Information Sources Using Context Logic. In *Proceedings of AAAI Spring Symposium on Information Gathering from Distributed Heterogeneous Environments*.

Fisher, M. and C. Ghidini. 1999. Programming Resource-Bounded Deliberative Agents. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI'99)*, pages 200–206. Morgan Kaufmann Publ., Inc.

Ghidini, C. 1999. Modelling (Un)Bounded Beliefs. In P. Bouquet, L. Serafini, P. Brezillon, M. Benerecetti, and F. Castellani, eds., *Modelling and Using Context – Proceedings of the 2nd International and Interdisciplinary Conference, Context'99*, vol. 1688 of *Lecture Notes in Artificial Intelligence*, pages 145–158. Springer Verlag - Heidelberg.

Ghidini, C. and F. Giunchiglia. 2001. Local models semantics, or contextual reasoning = locality + compatibility. *Artificial Intelligence* 127(2):221–259.

Ghidini, C. and L. Serafini. 1998a. Information Integration for Electronic Commerce. In C. Sierra and P. Noriega, eds., *Proceedings of the Workshop on Agent Mediated Electronic Trading (AMET'98)*, vol. 1571 of *LNAI*, pages 189–206. Minneapolis, USA: Springer Verlag.

Ghidini, C. and L. Serafini. 1998b. Model Theoretic Semantics for Information Integration. In F. Giunchiglia, ed., *Proceedings AIMSA'98, 8th International Conference on Artificial Intelligence, Methodology, Systems, and Applications*, vol. 1480 of *LNAI*, pages 267–280. Sozopol, Bulgaria: Springer Verlag.

Giunchiglia, E. and F. Giunchiglia. 1996. Ideal and Real Belief about Belief. In *Practical Reasoning, International Conference on Formal and Applied Practical Reasoning, FAPR'96*, no. 1085 in Lecture Notes in Artificial Intelligence, pages 261–275. Springer Verlag.

Giunchiglia, F. 1993. Contextual reasoning. *Epistemologia, special issue on I Linguaggi e le Macchine* XVI:345–364. Short version in Proceedings IJCAI'93 Workshop on Using Knowledge in its Context, Chambery, France, 1993, pp. 39–49.

Giunchiglia, F. and L. Serafini. 1994. Multilanguage hierarchical logics, or: how we can do without modal logics. *Artificial Intelligence* 65(1):29–70.

Giunchiglia, F., L. Serafini, E. Giunchiglia, and M. Frixione. 1993. Non-Omniscient Belief as Context-Based Reasoning. In *Proc. of the 13th International Joint Conference on Artificial Intelligence*, pages 548–554. Chambery, France.

Guha, R.V. 1991. *Contexts: a Formalization and some Applications*. Ph.D. thesis, Department of Computer Science, Stanford University. Also report No. STAN-CS-91-1399-Thesis.

Lenat, D. B. 1995. Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM* 38(11):33–38.

McCarthy, J. 1987. Generality in Artificial Intelligence. *Communications of ACM* 30(12):1030–1035. Also in V. Lifschitz (ed.), *Formalizing common sense: papers by John McCarthy*, Ablex Publ., 1990, pp. 226–236.

McCarthy, J. 1993. Notes on Formalizing Context. In *Proc. of the 13th International Joint Conference on Artificial Intelligence*, pages 555–560. Chambery, France.

Mylopoulos, J. and R. Motschnig-Pitrik. 1995. Partitioning Information Bases with Contexts. In *Third International Conference on Cooperative Information Systems*. Vienna.

Noriega, P. and C. Sierra. 1996. Towards Layered Dialogical Agents. In J. P. Müller, M. J. Wooldridge, and N. R. Jennings, eds., *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages (ATAL-96), ECAI-96*, vol. 1193 of *LNAI*, pages 173–189. Budapest, Hungary.

Parsons, S., C. Sierra, and N. R. Jennings. 1998. Agents that reason and negotiate by arguing. *Journal of Logic and Computation* 8(3):261–292.

Theodorakis, M., A. Analyti, P. Constantopoulos, and N. Spyratos. 1998. Context in Information Bases. In *Proceedings of the 3rd International Conference on Cooperative Information Systems (CoopIS'98)*, pages 260–270. New York City, NY, USA: IEEE Computer Society.

3

Contextual Intensional Logic: Type-Theoretic and Dynamic Considerations

Richmond H. Thomason

3.1 Introduction

This chapter presents Contextual Intensional Logic, a type-theoretic logic intended as a general foundation for reasoning about context. I motivate and illustrate the logical framework, and discuss a dynamic version that provides for context-changing operators.

In several previous works, Thomason (1997, 1998), I proposed and explored the idea of formulating contextual logic as a version of type theory. Here I improve, refine and extend the previous presentations (which were altogether too sketchy) of these type-theoretic ideas.

There are a number of logical advantages to type theory, which make it an appropriate and fruitful starting-point for a formalization of context.

(1) The underlying logical architecture, which goes back to Church (1940), is beautifully simple and has been thoroughly investigated by logicians.
(2) The framework of types provides a rich, highly structured ontology that is potentially useful in formalization.
(3) The use of types provides conceptual clarity.
(4) An appropriate home for context can be found in the types of a natural extension of Richard Montague's Intensional Logic (IL), Montague (1970), Gallin (1975), Anderson (1984).

(5) IL has has been the dominant formalism for the logical inter-
pretation of natural language. Using it provides direct con-
nections to an extensive body of work in natural language
semantics. So the type theoretic approach situates the the-
ory of context at once in a broader, well-explored approach
to the semantics of human languages.

In this paper, I will introduce and explain the basic ideas of the
formalism, using these to illustrate and support point (4), above. I
will also discuss some ways in which the formalism can (and should) be
extended, with more detailed remarks concerning dynamic extensions of
the logic than those that appeared in Thomason (1999), the conference
publication of this paper.

3.2   Brief Introduction to IL

Any version of type theory will involve—in addition to a domain of
individuals, with variables ranging over this domain—domains that
correspond to higher-order types: sets of individuals, sets of sets of
individuals, etc. The formalizations of type theory based on Church
(1940) use functional (i.e., lambda) abstraction to organize the higher-
order domains. In general, where D_1 and D_2 are domains, the set $D_2^{D_1}$
of functions from D_1 to D_2 is also a domain.

This leads to the following recursive definition of types, in which the
types for individuals and truth values are primitive, and all other types
are functional.

(2.1) e is a type.

(2.2) t is a type.

(2.3) If σ and τ are types, so is $\langle \sigma, \tau \rangle$.

Here, e is the type of individuals (*e*ntities), t is the type of truth-values,
and $\langle \sigma, \tau \rangle$ stands for the type of functions from objects of type σ to
objects of type τ.

The language of type theory has an infinite set of variables of each
type.[1] The language has only three primitive syntactic constructions.

Identity: If ζ and ξ are expressions of type τ, so is
                              $\zeta = \xi$.

Functional application: If ζ is an expression of type $\langle \sigma, \tau \rangle$ and ξ
                              is an expression of type σ, then $\zeta(\xi)$ is an
                              expression of type τ.

---

[1]In writing formulas of intensional logic, I will label the first occurrence of a
variable with its type unless the type is e, in which case the label may be omitted.
Later occurrences will not be marked for type; no confusion can arise, as long as all
independent uses of bound variables involve distinct variables.

Lambda abstraction: If ζ is an expression of type τ, then $\lambda x_\sigma \zeta$ is an expression of type $\langle \sigma, \tau \rangle$.

With these resources, the full set of boolean operations can be defined, as well as universal and existential quantification over domains of any type. The model theory of the logic is straightforward; arbitrary domains are assigned to primitive types, the domain of $\langle \sigma, \tau \rangle$ is the set of functions from the domain of type σ to the domain of type τ, $=$ is interpreted as identity, () is interpreted as functional application, λ is interpreted as functional abstraction. See Gallin (1975) for details on these matters.

A number of ontological policies come along with this approach to types. In particular, sets are represented by the corresponding characteristic functions. That is, a set of objects of type τ is represented as a function of type $\langle \tau, t \rangle$, and n-place functions from types $\langle \sigma_1, \sigma_2, \ldots, \sigma_n \rangle$ to type τ are represented as a nested type $\langle \sigma_1, \langle \sigma_2, \ldots \langle \sigma_n, \tau \rangle \ldots \rangle \rangle$ consisting of iterations of 1-place functions. According to these policies, for instance, a 2-place function from individuals to individuals would have type $\langle e, \langle e, e \rangle \rangle$; a function of this type inputs an individual and outputs a function from individuals to individuals. A 2-place relation between individuals and sets of individuals would have type $\langle e, \langle \langle e, t \rangle, t \rangle \rangle$. An object of this type would be a function that inputs a first-order object and outputs a function that inputs a set of first-order objects (which itself is a function from first-order objects to truth-values) and outputs a truth-value.

Church's formalism uses the domain of truth values to interpret sentences. Obviously, a domain containing only two values will be unable to represent sentence meanings adequately, an inadequacy that is reflected in the inability of the logic to deal with *propositional attitudes* such as belief. Montague's IL seeks to remedy this problem by introducing a third primitive type w, the type of possible worlds. This makes available a type $\tau\text{-Prop} = \langle w, t \rangle$ of propositions.[2]

In natural language, we apparently do not refer explicitly to possible worlds; similarly, the syntax of IL gives worlds an implicit role. There are no constants or variables of type w, so there is no explicit lambda abstraction over the domain of worlds. Instead, IL has operators that form intensions and extensions.

Intension: If ζ is an expression of type τ, then $^\wedge \zeta$ is an expression of type $\langle w, \tau \rangle$.

---

[2]Actually, Montague's type apparatus extends the type-forming operators rather than the primitive types. But it is equivalent to the version I present here; see Gallin (1975) for details.

Extension: If ζ is an expression of type $\langle w, \tau \rangle$, then $^{\vee}\zeta$ is an expression of type τ.

To illustrate the typical formalizations that go along with IL, take an example like

(2.4) Some guest wants to stay and eat.

The noun phrase 'some guest' can be thought of as saying that the (characteristic function) of the intersection of the set of guests with a certain set—the set of things that want to stay and eat—is nonempty. Thus, 'some guest' inputs something of type $\langle e, t \rangle$ and returns a truth value. Therefore it has type $\langle \langle e, t \rangle, t \rangle$. Again, since 'guest' has type $\langle e, t \rangle$, 'some' must have type $\langle \langle e, t \rangle, \langle \langle e, t \rangle, t \rangle \rangle$. The verb 'want' relates a property—the intension of a set—and an individual. So we can think of 'want' as inputing the intension of a set, and returning a function that inputs an individual and returns a truth value. Therefore, 'want' has type $\langle \langle w, \langle e, t \rangle \rangle, \langle e, t \rangle \rangle$.

This yields the following formalization of (2.4).

Type assignments: some : $\langle \langle e, t \rangle, t \rangle$
                       guest : $\langle e, t \rangle$
                       want : $\langle \langle w, \langle e, t \rangle \rangle, \langle e, t \rangle \rangle$
                       stay : $\langle e, t \rangle$
                       eat : $\langle e, t \rangle$

Formalization: some(guest)($\lambda x[$want($\lambda y[$stay(y) $\wedge$ eat(y)])(x)])

This rather natural and modular correspondence between the English syntax and the logic can be deployed over a wide range of languages and constructions. Over the last thirty years it has become the dominant approach to the semantics of natural language. See, for instance, Partee and Hendriks (1996) for further information and references.

It is time to be more explicit about the model theory. A *frame* $\mathcal{F}$ assigns a nonempty set $\mathrm{Dom}_{\mathcal{F}}(\tau)$ to each primitive type $\tau \in \{e, w\}$; $\mathrm{Dom}_{\mathcal{F}}(t) = \{\top, \bot\}$, where $\top$ and $\bot$ are two arbitrary, fixed objects. The domains assigned to complex types in $\mathcal{F}$ are then defined by the following equation.

$$\mathrm{Dom}_{\mathcal{F}}(\langle \sigma, \tau \rangle) = \mathrm{Dom}_{\mathcal{F}}(\tau)^{\mathrm{Dom}_{\mathcal{F}}(\sigma)}.$$

A model $\mathcal{M}$ on $\mathcal{F}$ assigns each constant α of type τ a value $[\![\alpha]\!]_{\mathcal{M}, \mathrm{w}}$ in $\mathrm{Dom}_{\mathcal{M}}(\tau) = \mathrm{Dom}_{\mathcal{F}}(\tau)$, for each world $\mathrm{w} \in \mathrm{Dom}_{\mathcal{M}}(w)$; $\mathcal{M}$ assigns each variable x of type τ a value $[\![x]\!]_{\mathcal{M}}$ in $\mathrm{Dom}_{\mathcal{M}}(\tau)$; we stipulate that for all w, $[\![x]\!]_{\mathcal{M}, \mathrm{w}} = [\![x]\!]_{\mathcal{M}}$.

The values of complex expressions are defined as follows, where $\mathcal{M}^{\mathrm{d}}/x$ is the model like $\mathcal{M}$ except that $[\![x]\!]_{\mathcal{M}^{\mathrm{d}}/x} = \mathrm{d}$.

Where ζ and ξ are expressions of type τ, $[\![\zeta = \xi]\!]_{\mathcal{M},w} = \top$ iff $[\![\zeta]\!]_{\mathcal{M},w} = [\![\xi]\!]_{\mathcal{M},w}$.

Where ζ is an expression of type $\langle \sigma, \tau \rangle$ and ξ is an expression of type σ, $[\![\zeta(\xi)]\!]_{\mathcal{M},w} = [\![\zeta]\!]_{\mathcal{M},w}([\![\xi]\!]_{\mathcal{M},w})$.

Where ζ is an expression of type τ, $[\![\lambda x_\sigma(\zeta)]\!]_{\mathcal{M},w} = f$, where f is the function from $\mathrm{Dom}_{\mathcal{M}}(\sigma)$ to $\mathrm{Dom}_{\mathcal{M}}(\tau)$ such that $f(d) = [\![\zeta]\!]_{\mathcal{M}^{d/x},w}$.

Where ζ is an expression of type τ, $[\![^\wedge\zeta]\!]_{\mathcal{M},w} = f$, where f is the function from $\mathrm{Dom}_{\mathcal{M}}(w)$ to $\mathrm{Dom}_{\mathcal{M}}(\tau)$ such that $f(w') = [\![\zeta]\!]_{\mathcal{M},w'}$.

Where ζ is an expression of type $\langle w, \tau \rangle$, $[\![^\vee\zeta]\!]_{\mathcal{M},w} = [\![\zeta]\!]_{\mathcal{M},w}(w)$.

Treating propositions as sets of possible worlds is, of course, problematic. (See, for instance, Moses (1988), Giunchiglia et al. (1993).) But despite its foundational problems, this approach has been pursued with some success in philosophical logic, computer science, economics, and natural language semantics. It is certainly possible to generalize the possible worlds approach to intensionality to obtain a less restrictive account of propositions. But it seems to me that such generalizations are premature. Without constraints, hyperintensional theories are uninformative. Appropriate constraints seem to require better general models of the reasoning agent than we have at present. Until such models are developed, I prefer to use the possible worlds formalisms, which in any case have many features that would need to be preserved in any more general approach.

Richard Montague showed that the framework of intensional logic provides a type for propositions, and that many other types that are ontologically natural and useful in the interpretation of natural language semantics can be characterized in IL. I will now explain why this framework also provides an appropriate and useful architecture for thinking about contexts.

3.3 Contexts as Modalities

The integration of knowledge sources and modules seems at present to be the origin of the most detailed and illuminating examples of ways in which the theory of context can be used.[3] If we look at a context as a knowledge source, the simplest way to model contexts in a type theoretic setting would be to identify a context with the set of propositions that it delivers. This would locate contexts in the type $\langle \langle w, t \rangle, t \rangle$, which is the type that I assigned to contexts in Thomason (1998).

---

[3]See, especially, the example in (McCarthy and Buvač, 1995, Section 6).

48 / Richmond H. Thomason

This type-theoretic account of contexts would suffice for applications in which an agent (which is not itself a context) is simultaneously accessing information from many contexts. Each context would then deliver its set of propositions to the collecting agent. But if we want to allow *contexts* to access information from other contexts, we need an enriched representation of contexts.

The problem is this. The type $\langle\langle w, t\rangle, t\rangle$ makes contexts have propositions as inputs. This intensionality of contexts is mandated by the desired applications: we certainly do not want a context to support *every* true sentence if it supports *any* true sentence. Since the output assigned to a proposition by a context c is only a truth value according to this account, although we can access a proposition p that holds in c there will be no way to construct from c and p the proposition that says that p holds in c. Therefore, we can't pass the output of applying a context to a proposition on to another context as an input.

To solve this problem, we need to enrich the output of the type that is associated with contexts. Iteration of contexts can be managed in several ways within intensional logic; the following formalization corresponds directly to multi-agent epistemic logics, of the sort discussed in Fagin et al. (1995), in which a binary relation R_a over worlds is associated with an epistemic agent a.

This standard relational semantics for modal operators specifies that $\square_a \phi$ is true in w if and only if ϕ is true in all worlds w′ such that $wR_a w'$. The idea can be captured in IL by (i) locating contexts in the type $\tau\text{-Mod} = \langle\langle w, t\rangle, \langle w, t\rangle\rangle = \langle\tau\text{-Prop}, \tau\text{-Prop}\rangle$ and (ii) for each context M adding following axiom, which guarantees that the behavior of the context is determined according to the standard constraint on modalities.[4]

(3.1) $\forall p_{\tau\text{-Prop}}[^\vee[M(^\wedge[^\vee p \to {}^\vee q])] \to [^\vee M(p) \to {}^\vee M(q)]]$

Here, M is a constant of type $\tau\text{-Mod}$.

Within this framework we can provide a definition of McCarthy's *ist* relation, reconstructed here as a relation between contexts and the *propositions* (not the sentences) that hold in these contexts.

(3.2) $ist_1 = \lambda x_{\tau\text{-Mod}} \lambda y_{\tau\text{-Prop}} x(y)$

This definition gives *ist* the type $\langle\tau\text{-Mod}, \langle\tau\text{-Prop}, \tau\text{-Prop}\rangle\rangle$, which inputs a modality and a proposition, and outputs a proposition.

This approach to context has several shortcomings, some of them substantive and some of them a matter of public relations. I will address the less important class of problems first, leaving the substantive issues

---

[4]Of course, we can add additional constraints on the relation R if we wish.

for the remaining sections.

If we treat contexts as modal operators, it is hard to see what is new about the logic of context. Modal logic is a well developed area of logic that has received a great deal of attention over the last forty years. Regarding contextual logic as a branch of modal logic seems to leave relatively little work for us to do.

While it is true that this conservative approach may rule out a more logically creative program, it still leaves room for some innovations, because (as I will argue in the next section), the logic of context can't in fact be identified with modal logic. But treating it as a generalization which preserves the main features of modal logic has many advantages. First, it enables us to import results and applications from modal logic. It is in fact, very useful to regard contexts as simple epistemic agents, agents which know information about other agents and can communicate with other, similar agents. This makes ideas concerning protocol design and knowledge-based programming available in the logic of context.[5] We may also be able to import the techniques that have been developed for modal theorem proving. (See Stone (1998).)

3.4   Contextual Intensional Logic

There is a serious limitation to this approach to context; it will not deal with cases in which the meanings of terms can differ from context to context. The computational literature has regarded variation in meaning as an essential application of the logic of context. But modal logic can't track shifts of meaning among the various senses of ambiguous expressions. Take the simplest case: two "personal assistant" databases with two users, a and b. Both databases record information about their users, using an internal constant @USER to refer to their users. To merge this information coherently, we have to assign different propositions to expressions like

(4.1) BIRTHDATE(@USER, $\langle 4, 4, 1969 \rangle$).

But modal logic has no natural way to represent the reasoning that produces these different assignments.[6]

Translating ideas from Kaplan (1978) into the type-theoretic framework, I propose to address this problem by introducing a fourth primitive type: the type i of indices. Kaplan thinks of indices as arguments

---

[5]See Fagin et al. (1995) for discussion of these matters, and for further references.

[6]We could, of course, formalize the reasoning as syntactic, i.e. we could treat it as reasoning about expressions. I will not explore this alternative here; despite its apparent naturalness, it is much less satisfactory in the long run, I believe, than the intensional approach that I assume here. The main formal problem that a syntactic approach raises, of course, is that it reintroduces the semantic paradoxes.

that determine the interpretations of "indexical expressions" such as 'I', 'here', and 'now'.

Indexicals are certainly a dramatic illustration of context-dependence, but typical uses of contextual reasoning to integrate knowledge sources also need to be able to deal with the ambiguity of lexical expressions that are not usually thought of as indexical; for instance, a case in which one database used by a bank uses 'Account' to refer to active banking accounts, while another database used by the same bank uses 'Account' to refer to active and inactive banking accounts, while still a third database uses 'Account' to refer to brokerage accounts.

I want to extend the range of Kaplan's indices, by thinking of an index as a simultaneous disambiguation of all the relevant lexical indexicalities and ambiguities that can arise in an application. Call such a simultaneous disambiguation a *contextualization*. If the only contextualizations arise from 'I' and 'here', we can identify an index with a pair consisting of a person and a place. If the only contextualizations serve to resolve the ambiguity of ten lexical items, each of them having just two possible meanings, then we will need indices corresponding to the 2^{10} possible disambiguations.

As a first step in presenting Contextual Intensional Logic (CIL), we extend the types by introducing a primitive type i for indices. The recursive definition of complex types then provides for types such as $\langle i, \langle w, t \rangle \rangle$ (the type of a context-dependent proposition) and $\langle i, \langle w, \langle e, t \rangle \rangle \rangle$ (the type of a context-dependent property).

Expressions are now evaluated in a model $\mathcal{M}$ relative to an index and a world; $[\![\zeta]\!]_{\mathcal{M},\mathrm{i},\mathrm{w}}$ is the value assigned by model $\mathcal{M}$ to ζ at i and w.

We will provide operators $^\sqcap$ and $^\sqcup$, analogous to the corresponding operators used to manage intensionality in IL. The interpretation of expressions involving these operators is as follows.

Where ζ is an expression of type τ, $[\![^\sqcap\zeta]\!]_{\mathcal{M},\mathrm{i},\mathrm{w}} = \mathrm{f}$, where f is the function from $\mathrm{Dom}_{\mathcal{M}}(i)$ to $\mathrm{Dom}_{\mathcal{M}}(\tau)$ such that $\mathrm{f}(\mathrm{i'}) = [\![\zeta]\!]_{\mathcal{M},\mathrm{i'},\mathrm{w}}$.

Where ζ is an expression of type $\langle i, \tau \rangle$, $[\![^\sqcup\zeta]\!]_{\mathcal{M},\mathrm{i},\mathrm{w}} = [\![\zeta]\!]_{\mathcal{M},\mathrm{i},\mathrm{w}}(\mathrm{i})$.

The treatment of the type indices in CIL, however, is not entirely similar to that of worlds in Montague's IL. We will allow constants and variables of type i, and explicit lambda abstraction over these types. This is needed in order to provide a means of defining functions that return context-dependent values. Definition (4.3), below, of McCarthy's *ist* construction, is an example.

Assigning an extension to a contextualized expression involves disambiguation, which yields an intension, and evaluation, where the in-

tension is given a value relative to an index in a world. Thus, to interpret an expression like

(4.2) 'I'm over 21 years old'

we first need to identify the speaker, s; this yields the proposition that is true in a world if and only if s is over 21 years old in that world.

Following Kaplan, we'll refer to functions from indices to context-independent meanings as "characters." We can then represent the assignment of an extension to a linguistic expression as a three-stage process. First, the semantics of the language associates a character with an expression. The character of (4.2), for instance, is a potentiality to yield different propositions at different indices. This character is then evaluated at an index to yield a *content*, which finally may be evaluated in a world to produce an extension. In the case of sentence (4.2), whose only indexical is 'I', we can think of the character as a function from persons to propositions. This function inputs a person and returns the proposition that the person is over 21 years old. I will say that (4.2) *expresses* this character. Evaluating this character at a person s produces a function from possible worlds to truth values that returns $\top$ on input w if and only if s is over 21 years old in w.

I will say that a character *indicates* its content at an index. So (4.2) expresses a character which at the index Ann[7] indicates the proposition that Ann is over 21 years old.

How do my modalities and indices relate to McCarthy's contexts? As I see it, contexts of the sort envisaged by McCarthy perform two different functions: they disambiguate and they serve as knowledge sources. From the standpoint of CIL these roles are separated: the first is represented by an index, or object of type i; the second is represented by a modality, or object of type τ-Mod. *Therefore, a context is a pair consisting of an index and a modality.*

This idea yields no very natural type for contexts in CIL, because CIL provides no direct encoding for ordered pairs. If we use one of the (unnatural) encodings in CIL of the cross product $\sigma \times \tau$ of σ with τ, or (better) if we add a cross-product operation to the underlying type definition, we can situate *indexical contexts* in the type τ_2-Con $= i \times \tau_1$-Con.

But we can avoid the need to provide a type for CIL contexts by separating indices and modalities as arguments of *ist*, treating *ist*, say, as a function from objects of type i to functions from objects of type τ-Mod to appropriate values. If we want this *ist* to iterate, it should output a

---

[7]For the moment I'm identifying indices with persons; this does no harm if the first person pronoun is the only indixical expression.

propositional character. The most general account of *ist*, then, would give it the type $\langle i, \langle \tau\text{-Mod}, \langle \tau\text{-Char-Prop}, \tau\text{-Char-Prop}\rangle\rangle\rangle$. This *ist* inputs a modality M and an index i. Then, given a propositional character p, it outputs a propositional character q. At any index j, q indicates the proposition that p is necessary according to M, where p is the proposition indicated by q at i. Therefore, expressions of the form $ist(\eta)(\zeta)(\phi)$ designate constant characters, ones that express the same proposition at each index. So we can in fact assign the simpler type

$$\langle i, \langle \tau\text{-Mod}, \langle \tau\text{-Char-Prop}, \tau\text{-Prop}\rangle\rangle\rangle$$

to *ist*. Accordingly, *ist* outputs a proposition rather than a propositional character. To iterate this *ist*, we use the $\ulcorner$ operator, as follows:

$$ist(\eta)(\zeta)(\ulcorner ist(\eta')(\zeta')(\phi)).$$

This simplified *ist* can be defined as follows using lambda abstraction.

(4.3) $ist \; = \; \lambda u_i \lambda x_{\tau\text{-Mod}} \lambda p_{\tau\text{-Char-Prop}} \, x(p(u))$

According to this definition, given an index i and a modality M, *ist* inputs a propositional character p and outputs the proposition that the proposition expressed by p at i is necessary according to M. (4.3) represents the final, official account of *ist* that I wish to recommend.

We will use the alternative notation $ist(\eta, \zeta, \phi)$ for $ist(\eta)(\zeta)(\phi)$. From (4.3), we obtain the following logical equivalence.

(Ist Conversion) $ist(\eta)(\zeta)(\phi)$ and $\zeta(\phi(\eta))$ are equivalent.

These ideas are contrary in spirit to remarks in which McCarthy suggests that contexts should be formally treated as primitives. I take McCarthy's remarks to mean that, although in some applications we can axiomatize general knowledge about contexts, it is pointless to attempt to define contexts. Actually, I agree with McCarthy that relatively little of the work that needs to be done to explicate contexts can be done with definitions. But I do think it is enlightening and helpful to separate contexts into two components, one of them (the index) dealing with indexicality and ambiguity, while the other (the modality) deals with knowledge.

To show how the intended sort of applications would be formalized in CIL, it will be helpful to present a simple example. A group of people including Ann, Bob and Charlie use personal databases to manage their calendars. The language of the databases has a first-person pronoun referring to the database user, as well as constants referring to Ann, Bob and Charlie. In Ann's database, 'I' refers to Ann. Assume that the databases contain information about meetings, in the form of a set

of triples whose whose first and second members denote people, and whose third member denotes a time. They also contain information about databases, in the form of a set of quadruples whose first member denotes a database, whose second and third members are names of people, and whose fourth member denotes a time. In this example, we do not distinguish between people and their databases; a constant a of type e refers to Ann (and to her database); similarly, b refers to Bob and c to Charlie.

The databases and the information they contain can be formalized at three levels. At the *database format level*, the representation is close to the one actually manipulated by the databases. The following is a format level description of Ann's database.

Database format level:

> (ADB1.1) $\langle I, b, 9 \rangle$
> (ADB1.2) $\langle I, c, 10 \rangle$
> (ADB1.3) $\langle b, I, a, 9 \rangle$

I'm supposing that the last entry is a direct readout of the contents of Bob's database; so in (ADB1.3), 'I' refers to Bob and the entry represents the same 9 o'clock appointment that is entered in (ADB1.1).

The database format level representations use a single indexical, 'I'; the ways in which the content of this indexical vary would have to be taken into account in writing the database procedures. For instance, in checking whether Bob's database is consistent with Ann's, Ann's database would have to convert (ADB1.3) to $\langle b, I, 9 \rangle$ before comparing it with the triples representing Ann's appointments.

At the *knowledge level*,[8] modal operators are used to encode the knowledge ascribed to different sources. But, as in ordinary modal logic, there is no explicit apparatus for keeping track of indexicals, so all indexicals that occur at the database format level have to be replaced by equivalent, nonindexical terms.

In the following knowledge level description of Ann's database, `[a]` and `[b]` are constants of type τ-Mod, representing the modalities corresponding to Ann's and Bob's databases.[9]

Knowledge level:

> (ADB2.1) `[a]`MEET$(a, b, 9)$
> (ADB2.2) `[a]`MEET$(a, c, 10)$
> (ADB2.3) `[a]``[b]`MEET$(b, a, 9)$

---

[8]I hope the use of Allen Newell's term in a slightly different sense will cause no confusion.

[9]MEET has type $\langle e, \langle e, \langle e, t \rangle \rangle \rangle$. MEET$(\eta, \zeta, \xi)$ is an abbreviation of $[[\text{MEET}(\eta)](\zeta)](\xi)$.

These axioms correspond to the most natural way of talking about what Ann (or her database) knows: Ann knows that she is to meet with Bob at 9, Ann knows that she is to meet with Charlie at 10, Ann knows that Bob knows he is to meet with her at 9. In this simple example the required translation of the indexical representations into neutral terms is not especially difficult, but it can be a challenge in more complicated cases. Moreover, it is plausible to assume that (ADB2.3) holds because Ann's and Bob's databases are communicating. When this happens, Ann's database somehow has to impose the right interpretation on the presence of $\langle I, a, 9 \rangle$ in Bob's database; somehow, Ann's database has to know that 'I' refers to Bob when Bob uses it.

These are problems of the sort that McCarthy's *ist* is designed to solve. The somewhat different account of *ist* in CIL performs a similar function. First, in this example we identify indices with people; $\text{Dom}_{\mathcal{M}}(i) = \{a, b, c\}$, where $a, b, c \in \text{Dom}_{\mathcal{M}}(e)$. Now, nothing prevents domains of different types from overlapping, but each expression must have a unique type; so we need different constants for people considered as individuals and for people considered as indices. We will use a_e, b_e, and c_e for the former, and a_i, b_i, and c_i for the latter. Although a_e and a_i, for instance, both denote the same thing, $a_e = a_i$ is not a well-formed expression of CIL.

We then let I be a constant of type e such that $[\![I]\!]_{\mathcal{M}, i, w} = i$ for all $i \in \text{Dom}_{\mathcal{M}}(i)$; $^{\sqcap}I$ returns a in Ann's database, b in Bob's database, and c in Charlie's database. In other words, I is assigned the character intuitively associated with the first person singular pronoun.

Here is the *context level* formalization in CIL of Ann's database, using *ist*.

Context level:

(ADB3.1) $^{\vee}ist(a_i, \texttt{[a]}, ^{\sqcap\wedge}\text{Meet}(I, b_e, 9))$[10]

(ADB3.2) $^{\vee}ist(a_i, \texttt{[a]}, ^{\sqcap\wedge}\text{Meet}(I, c_e, 10))$

(ADB3.3) $^{\vee}ist(a_i, \texttt{[a]}, ^{\sqcap}ist(b_i, \texttt{[b]}, ^{\sqcap\wedge}\text{Meet}(I, a_e, 9)))$

Recall that the type of our simplified *ist* as defined in (4.3) is $\langle i, \langle \tau\text{-Mod}, \langle \tau\text{-Char-Prop}, \tau\text{-Prop}\rangle\rangle\rangle$. With this in mind, let's check formula (ADB3.1) for type coherence. The first argument of *ist*, a_i, has type i. The second argument, $\texttt{[a]}$, has type τ-Mod. The third and last argument, $^{\sqcap\wedge}\text{Meet}(I, a_e, 9)$, has type τ-Char-Prop. So the arguments have the required types. The formula $ist(a_i, \texttt{[a]}, ^{\sqcap\wedge}\text{Meet}(I, a_e, 9))$ therefore has type τ-Prop, so finally (ADB3.1) has type t and denotes a truth-value, as an axiom should.

Representing what these context-level formulas say in simple English can become difficult in moderately complex cases, because nat-

ural languages seem to resist fully explicit representations of context. For instance, $ist(a_i, [b], \ulcorner{}^\wedge\textsc{Meet}(b_e, I, 9))$ says that the propositional character expressed by 'Bob has a meeting with me at 9' indicates for Ann a proposition that Bob knows. And

$$ist(a_i, [a], ist(b_i, [b], {}^\wedge\textsc{Meet}(I, a_e, 9))$$

says that the propositional character expressed by (4.4) indicates for Ann a proposition that Ann knows.

(4.4) The propositional character expressed by (4.5) indicates for Bob a proposition that Bob knows.

(4.5) I have a meeting with Ann at 9.

Unlike the other two levels, the context level provides an explicit language in which protocols for communicating databases can be specified and proved correct. Although, of course, the need for such proofs is not compelling in this simple example, there is a genuine need for specifications and correctness proofs even in moderately complicated cases of this sort.

For example, consider how CIL can be used to formalize the process that leads from Ann's observation (ADB1.3) of an entry in Bob's database to the statement (ADB2.3) of the result of this observation on Ann's knowledge. The steps in this reasoning require the following domain axioms.

(DB.1) $\Box({}^\wedge\forall x \forall y[[\ulcorner{}^\wedge\textsc{Meet}(I, x, y)](b_i) = {}^\wedge\textsc{Meet}(b_e, x, y)])$

(DB.2) $\Box({}^\wedge\forall u_i[\ulcorner[b]({}^\wedge\textsc{Meet}(a_e, b_e, 9))](u_i) = {}^\wedge[b]({}^\wedge\textsc{Meet}(a_e, b_e, 9))])$

(DB.3) $\Box({}^\wedge\forall x \forall y \forall z[\textsc{Meet}(x, y, z) = \textsc{Meet}(y, x, z)])$

Axiom (DB.1) is one of a family of axioms that constrain the characters expressed by literals involving the first person pronoun; in this case, it implies that what 'I have a meeting with Ann at 9' expresses for Bob is the same proposition as that expressed (for any agent) by 'Bob has a meeting with Ann at 9'. Axiom (DB.2) is one of a family of axioms that assign constant characters to expressions that do not contain indexical terms.[11] The necessity sign in (DB.1) and (DB.2) is universal; so the axioms say that all databases know these equivalences. The first two axioms are analogous to McCarthy's "lifting axioms;" they allow the reformulation of index-dependent information across indices. Axiom (DB.3) states the symmetry of Meet in the first and second arguments. Again, the use of the necessity operator in the axiom means

---

[11]I will not explore these axioms here, or look into the question of how they might be derived from simpler principles.

that all databases know about this symmetry.

The formalization of the reasoning is as follows.

(i) Ann's database observes the triple $\langle I, a, 9 \rangle$ in Bob's database.

(ii) This observation is formalized at the context level as

$$^{\vee}ist(a_i, [\![a]\!], ^{\sqcap}ist(b_i, [\![b]\!], ^{\sqcap\wedge}\text{Meet}(I, a_e, 9)\,)\,).$$

Note that there is a simple, straightforward syntactic mapping between the terms of the observed triple and those of the formalization. This relationship is made possible by the use of characters.

(iii) We then obtain

$$^{\vee}ist(a_i, [\![a]\!], ^{\sqcap}[\![b]\!]([^{\sqcap\wedge}\text{Meet}(I, a_e, 9)](b_i)\,)\,)$$

from (ii) by **Ist Conversion**.

(iv) From (iii), we obtain

$$^{\vee}ist(a_i, [\![a]\!], ^{\sqcap}[[\![b]\!](^{\wedge}\text{Meet}(b_e, a_e, 9))]\,)$$

using (DB.1).

(v) Another application of **Ist Conversion** yields

$$^{\vee}[\![a]\!](^{\sqcap}[[\![b]\!](^{\wedge}\text{Meet}(b_e, a_e, 9))](a_i)\,).$$

(vi) Finally, applying (DB.3) to (vi) yields

$$^{\vee}[\![a]\!]([\![b]\!](^{\wedge}\text{Meet}(a_e, b_e, 9))\,).$$

This is the formalization in CIL of the knowledge level axiom (ADB2.3).

I hope that this example, simple as it is, gives a sense of how CIL could be useful in specifying inter-contextual reasoning.

3.5   Getting Dynamic

In McCarthy and Buvač (1995, 1998), McCarthy and Buvač suggest that operations of entering and exiting a named context would be useful additions to the logic of context, adding the thoughts that these operations should be like pushing and popping a stack, and that there should be analogies between contexts and the subproofs of the Fitch-style natural deduction format (Fitch (1952)). Similar remarks can be found in McCarthy (1993). The presentations that I have seen of these ideas are intriguing, but somewhat sketchy.

Ideas from dynamic logic (see, for instance, Harel (1984)) provide natural and principled methods for extending a purely declarative logic like CIL to one that contains dynamic operators. One advantage of this approach is that it provides a model theoretic interpretation of the resulting semantics.

In this section, I will indicate in a preliminary way how these techniques can be applied to CIL to obtain a version of Dynamic Contextual Intensional Logic (DCIL) with operators of the sort envisaged by McCarthy and Buvač.

A dynamic version of Intensional Logic is presented in Groenendijk and Stokhof (1990). This project was motivated by the desire to account for anaphora in natural language using dynamic variable binding. The requirements of dynamic contextual reasoning are rather different, and the logic DCIL presented below does not share many points of similarity with Groenendijk and Stokhof's Dynamic Montague Grammar, other than a common starting-point in an intensional type theory.

Recall that in CIL an expression ζ is assigned a value $[\![\zeta]\!]_{\mathcal{M},i,w}$ in a model $\mathcal{M}$; this value depends on a world w and an index i. We now revise this interpretation by adding a modality M in $\mathrm{Dom}_{\mathcal{M}}(\tau\text{-Mod})$ as another parameter, so that ζ is now assigned $[\![\zeta]\!]_{\mathcal{M},M,i,w}$. Constants and complex expressions are interpreted as before, so the interpretation of expressions does not depend on the new parameter. But since the two parameters M and i constitute a context, we can now say that in CIL expressions are interpreted relative to a context and a world. In fact, where $c = \langle M, i \rangle$, $M \in \mathrm{Dom}_{\mathcal{M}}(\tau\text{-Mod})$ and $i \in \mathrm{Dom}_{\mathcal{M}}(i)$, we can let $[\![\zeta]\!]_{\mathcal{M},M,i,w} = [\![\zeta]\!]_{\mathcal{M},c,w}$.

To put it crudely, the difference between a static and a dynamic model theory is that the latter relationalizes some of the parameters on which interpretation depends in the static case. Here, we are interested in a logic supporting dynamic operators that change contexts to contexts; so the dynamic models should involve relations over contexts. This can be accomplished by relativizing evaluation to a pair of contexts: $[\![\zeta]\!]_{\mathcal{M},c,c',w}$ is the value assigned to ζ relative to a world w and a pair of contexts c and c'. The semantic definitions of CIL for complex expressions are generalized unchanged to the dynamic case, except for the clause for identity. This is restricted so as to render identities static; that is, identities are not allowed to change the context. (See the discussion of static expressions, below.)

Where ζ and ξ are expressions of type τ, $[\![\zeta = \xi]\!]_{\mathcal{M},c,c',w} = \top$ iff $c = c'$ and $[\![\zeta]\!]_{\mathcal{M},c,c',w} = [\![\xi]\!]_{\mathcal{M},c,c',w}$.

Where ζ is an expression of type τ, $[\![\lambda x_\sigma(\zeta)]\!]_{\mathcal{M},c,c',w} = f$, where f is the function from $\mathrm{Dom}_{\mathcal{M}}(\sigma)$ to $\mathrm{Dom}_{\mathcal{M}}(\tau)$ such that $f(d) = [\![\zeta]\!]_{\mathcal{M}^{d/x},c,c'w}$.

Where ζ is an expression of type $\langle \sigma, \tau \rangle$ and ξ is an expression of type σ, $[\![\zeta(\xi)]\!]_{\mathcal{M},c,c',w} = [\![\zeta]\!]_{\mathcal{M},c,c',w}([\![\xi]\!]_{\mathcal{M},c,c',w})$.

Where ζ is an expression of type τ, $[\![^\wedge\zeta]\!]_{\mathcal{M},c,c',w} = f$, where f is the function from $\mathrm{Dom}_{\mathcal{M}}(w)$ to $\mathrm{Dom}_{\mathcal{M}}(\tau)$ such that $f(w') = [\![\zeta]\!]_{M,c,c',w'}$.

Where ζ is an expression of type $\langle w, \tau \rangle$, $[\![^\vee\zeta]\!]_{\mathcal{M},c,c',w} = [\![\zeta]\!]_{\mathcal{M},c,c',w}(w)$.

Where ζ is an expression of type τ and $c = \langle i, M \rangle$, $[\![^\sqcap\zeta]\!]_{\mathcal{M},c,c',w} = f$, where f is the function from $\mathrm{Dom}_{\mathcal{M}}(i)$ to $\mathrm{Dom}_{\mathcal{M}}(\tau)$ such that if $c = c'$ then $f(i') = [\![\zeta]\!]_{\mathcal{M},c'',c'',w}$ for $c'' = \langle i', M \rangle$, and if $c \neq c'$ then $f = \perp_{\mathcal{M},\langle i,\tau \rangle}$. ($\perp_{\mathcal{M}}$ is defined below.)

Where ζ is an expression of type $\langle i, \tau \rangle$ and $c = \langle i, M \rangle$, $[\![^\sqcup\zeta]\!]_{\mathcal{M},c,c',w} = [\![\zeta]\!]_{\mathcal{M},c,c',w}(i)$.

To provide a logical environment in which sentences—expressions of type t—can perform changes in the context, we have generalized evaluation to make it depend on a pair of contexts. For uniformity, we have had to treat the interpretation of expressions of arbitrary type in this way, but this generalization doesn't have a natural interpretation for many types; as far as I can see, for instance, although $[\![c_{\langle e,e \rangle}]\!]_{\mathcal{M},c,c',w}$ is formally defined, it has no clear intuitive meaning. We need to devise a way of making such cases vacuous. And we must do this in the context of a logic that is not partial, in which $[\![\zeta]\!]_{\mathcal{M},c,c',w}$ is always defined.[12]

Consider the type t; here we have some reliable dynamic intuitions. To say that a formula ϕ of type t is dynamically vacuous is to say that ϕ is a test; i.e., that it has no effect on context. Such formulas meet the following condition.

Definition 5.1. *Static type t expressions.*
An expression ϕ of type t is *static in $\mathcal{M}$* if for all c, c', w, if $c \neq c'$ then $[\![\phi]\!]_{\mathcal{M},c,c',w} = \perp$.

We generalize this definition to arbitrary types by choosing, for each type τ, a designated "null value" $\perp_{\mathcal{M},\tau}$. We make $\perp_{\mathcal{M},t} = \perp$; also, we want to ensure that $\perp_{\mathcal{M},\langle\sigma,\tau\rangle}(\perp_{\mathcal{M},\sigma}) = \perp_{\mathcal{M},\tau}$. We can do this by defining $\perp_{\mathcal{M},\tau}$ by the following induction on types.[13]

---

[12]Providing for partiality in CIL is a very natural step to take, but I believe it is better methodology to take these steps separately and incrementally. I have not yet tried to formulate a partial version of this logic.

[13]Note that the null values of type e, w, and i are chosen arbitrarily. Probably the most natural way to do this would be to create special elements for this purpose.

Definition 5.2. $\perp_{\mathcal{M},\tau}$.

> **Basis.** $\perp_{\mathcal{M},t} = \perp$; $\perp_{\mathcal{M},e} \in \mathrm{Dom}_{\mathcal{M}}(e)$; $\perp_{\mathcal{M},w} \in \mathrm{Dom}_{\mathcal{M}}(w)$;
> $\perp_{\mathcal{M},i} \in \mathrm{Dom}_{\mathcal{M}}(i)$.
> **Induction.** $\perp_{\mathcal{M},\langle\sigma,\tau\rangle}$ is the constant function f from $\mathrm{Dom}_{\mathcal{M}}(\sigma)$
> to $\mathrm{Dom}_{\mathcal{M}}(\tau)$ such that $\mathrm{f(d)} = \perp_{\mathcal{M},\tau}$ for all $\mathrm{d} \in \mathrm{Dom}_{\mathcal{M}}(\sigma)$.

An expression of type τ is static in case it receives the type τ null value relative to any pair of different contexts.

Definition 5.3. *Static type τ expressions.*
> An expression ξ of type τ is *static in* $\mathcal{M}$ if for all c, c′, w, if c $\neq$ c′
> then $[\![\xi]\!]_{\mathcal{M},c,c',w} = \perp_\tau$.

A static model is one that treats all constants as static. We can show that in a static model, all expressions are static.

Definition 5.4. *Static model.*
> A model $\mathcal{M}$ is *static* if for all constants α, α is static in $\mathcal{M}$.

Theorem 5.1. If $\mathcal{M}$ is static, then all expressions ξ are static in $\mathcal{M}$.

> *Proof.* By induction on the complexity of expressions.

There is a natural correspondence between models of CIL and static models of DCIL. To convert a CIL model $\mathcal{M}$ to an equivalent static DCIL model $\mathcal{M}'$, let $[\![\alpha]\!]_{\mathcal{M}',c,c,w} = [\![\alpha]\!]_{\mathcal{M},c,w}$ for all constants α. Use the same identity to convert a static DCIL model to an equivalent CIL model.

We can create special-purpose versions of DCIL with less trivial dynamic properties by adding truthlike constants that exhibit dynamic behavior. I will describe just one such example: ENTER.[14]

> ENTER is a constant of type $\langle\tau\text{-Mod}, \langle i, t\rangle\rangle$. $[\![\mathrm{ENTER}]\!]_{\mathcal{M},c,c',w}(\mathrm{M}, \mathrm{i}) =$
> $\top$ iff $c' = \langle M, i\rangle$, for all c, c′, M, i.

I believe that this approach is faithful, on the whole, to McCarthy and Buvač's ideas about context dynamics, and that it can be extended to formalize cases where information is extracted from a context by entering the context and performing inferential operations there. However, a dynamic "Exit" operator that returns to the previous context

---

[14]In the conference version of this chapter I discussed other dynamic constructions, and tried to indicate how they might be applied. That part of the paper was flawed; these matters require a more careful and extended treatment, something I hope to do in another paper.

cannot be formalized without making significant changes in the underlying model theory.

The usefulness of operators like ENTER($[b]$, b_i) is limited to cases in which explicit names of contexts are available. In some applications (for instance, ones where the contexts are specific information sources), these names may be available. In other cases, where contexts are implicit in a reasoning situation, they seem to be less directly accessible and we would have to find other operators to manage them.

Bear in mind that the material presented in this section is rather tentative. The formal properties of DCIL need to be explored more carefully; and I would very much like to develop sharper intuitions about the interactions between contextual dynamics, higher-order types, and intensionality.

3.6   Conclusion

There are several dimensions in which the logical framework that I have presented needs to be generalized in order to obtain adequate coverage:

(6.1) The logic needs to be made partial, to account for expressions which simply lack a value in some contexts.

(6.2) To account for default lifting rules, we need a nonmonotonic logic of context.

We have a general sense of what is involved in making a total logic partial, in making a static logic dynamic, and in making a monotonic logic nonmonotonic. For these reasons, I have adopted the strategy of concentrating on how to formulate an appropriate base logic to which these extensions can be made.

There are a number of approaches to the formalization of partial logics; indeed, the main problem with the logic of partiality, it seems to me, is that there are so many alternatives, and it is hard to select between them. Three-valued logic has been used in connection with the logic of context; see Buvač and Mason (1993). However, a four-valued logic is more symmetrical, and plausible arguments, starting with Belnap (1977), have been given for its computational usefulness. Most important for the project at hand, Muskens (1996) provides an extended study of how to modify IL using this approach to partiality. It is relatively straightforward to adopt Muskens' work to CIL.

There is, however, a much more ambitious application of partiality, according to which indices are regarded not as full, but as partial disambiguations of expressions. This program, which would require a more radical rethinking of the theory, may be needed to deal with applica-

tions of context to natural language interpretation, though perhaps it is unnecessary in cases in which indices correspond to carefully constructed knowledge sources. See, for instance, van Deemter and Peters (1996) for information on partial disambiguation.

As for nonmonotonicity, although circumscription is usually formulated in second-order extensional logic, it is relatively straightforward to add a theory of circumscription to IL. Similarly, I believe that other standard approaches to nonmonotonic logic could be adapted to a type theoretic framework. But this only provides a bare framework. The problems of developing a nonmonotonic logic capable of formalizing realistic problems in contextual reasoning would remain to be addressed.

None of these logical developments is entirely trivial, and in fact there is material here for many years of work. I hope to report on developments in these directions in the future.

Acknowledgments

I owe thanks to Chiara Ghidini for helpful comments on a draft of this chapter.

References

Anderson, C. Anthony. 1984. General intensional logic. In D. Gabbay and F. Guenther, eds., *Handbook of Philosophical Logic, Volume II: Extensions of Classical Logic*, pages 355–385. Dordrecht: D. Reidel Publishing Co.

Belnap, Nuel D. 1977. A useful four-valued logic. In J. Dunn and G. Epstein, eds., *Modern Uses of Multiple-Valued Logic*. Dordrecht: D. Reidel Publishing Co.

Buvač, Saša and Ian Mason. 1993. Propositional logic of context. In R. Fikes and W. Lehnert, eds., *Proceedings of the Eleventh National Conference on Artificial Intelligence*, pages 412–419. American Association for Artificial Intelligence, Menlo Park, California: AAAI Press.

Church, Alonzo. 1940. A formulation of the simple theory of types. *Journal of Symbolic Logic* 5:56–68.

Fagin, Ronald, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. 1995. *Reasoning About Knowledge*. Cambridge, Massachusetts: The MIT Press.

Fitch, Frederic B. 1952. *Symbolic Logic: An Introduction*. New York: The Ronald Press Co.

Gallin, Daniel. 1975. *Intensional and Higher-Order Logic*. Amsterdam: North-Holland Publishing Company.

Giunchiglia, Fausto, Luciano Serafini, Enrico Giunchiglia, and Marcello Frixione. 1993. Non-omniscient belief as context-based reasoning. In R. Bajcsy, ed., *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pages 548–554. San Mateo, California: Morgan Kaufmann.

Groenendijk, Jeroen and Martin Stokhof. 1990. Dynamic Montague grammar. Tech. Rep. LP–90–02, Institute for Language, Logic and Information, University of Amsterdam, Faculty of Mathematics and Computer Science, Roeterssraat 15, 1018WB Amsterdam, Holland.

Harel, David. 1984. Dynamic logic. In D. Gabbay and F. Guenther, eds., *Handbook of Philosophical Logic, Volume II: Extensions of Classical Logic*, vol. 2, pages 497–604. Dordrecht: D. Reidel Publishing Co.

Kaplan, David. 1978. On the logic of demonstratives. *Journal of Philosophical Logic* 8:81–98.

McCarthy, John. 1993. Notes on formalizing context. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pages 81–98. Los Altos, California: Morgan Kaufmann.

McCarthy, John and Saša Buvač. 1995. Formalizing context (expanded notes). Available from http://www-formal.stanford.edu/buvac.

McCarthy, John and Saša Buvač. 1998. Formalizing context (expanded notes). In A. Aliseda, R. van Glabbeek, and D. Westerståhl, eds., *Computing Natural Language*, pages 13–50. Stanford, California: CSLI Publications.

Montague, Richard. 1970. Pragmatics and intensional logic. *Synthèse* 22:68–94. Reprinted in *Formal Philosophy*, by Richard Montague, Yale University Press, New Haven, CT, 1974, pp. 119–147.

Moses, Yorav. 1988. Resource-bounded knowledge. In M. Y. Vardi, ed., *Theoretical Aspects of Reasoning About Knowledge: Proceedings of the Second Conference on Theoretical Aspects of Reasoning About Knowledge*, pages 261–276. San Francisco: Morgan Kaufmann.

Muskens, Reinhard. 1996. *Meaning and Partiality*. Cambridge, England: Cambridge University Press.

Partee, Barbara H. and Herman L.W. Hendriks. 1996. Montague grammar. In J. van Benthem and A. ter Meulen, eds., *Handbook of Logic and Language*, pages 5–91. Amsterdam: Elsevier Science Publishers.

Stone, Matthew. 1998. *Modality in Dialogue: Planning, Pragmatics and Computation*. Ph.D. dissertation, Computer Science Department, University of Pennsylvania, Philadelphia, Pennsylvania.

Thomason, Richmond H. 1997. Type theoretic foundations for context. In S. Buvač and Ł. Iwańska, eds., *Working Papers of the AAAI Fall Symposium on Context in Knowledge Representation and Natural Language*, pages 173–175. American Association for Artificial Intelligence, Menlo Park, California: American Association for Artificial Intelligence.

Thomason, Richmond H. 1998. Representing and reasoning with context. In J. Calmet and J. Plaza, eds., *Artificial Intelligence and Symbolic Computation; Proceedings of the International Conference AISC'98, Plattsburgh, New York*, pages 29–41. Berlin: Springer-Verlag.

Thomason, Richmond H. 1999. Type theoretic foundations for context, part 1: Contexts as complex type-theoretic objects. In P. Bouquet, L. Serafini,

P. Brézillon, M. Benerecetti, and F. Castellani, eds., *Modeling and Using Contexts: Proceedings of the Second International and Interdisciplinary Conference, CONTEXT'99*, pages 352–374. Berlin: Springer-Verlag.

van Deemter, Kees and Stanley Peters, eds. 1996. *Semantic Ambiguity and Underspecification*. Cambridge, England: Cambridge University Press.

March 22, 2006

4

The Search for the Semantic Grail

JOHN PERRY

4.1   Introduction

Consider Alphonse's explanation for thinking Sacramento is beautiful:

> Elwood saw a picture of Sacramento, which showed it to be quite
> beautiful, and so came to believe that it is a beautiful city. He told me
> that it was. I understood him, I trusted him, and that's how I came to
> believe that Sacramento is a beautiful city.

Here a familiar course of events is described. One person comes to
believe something. He tells someone else what he has come to believe.
That is, what he says is just what he believes. This person understands
him, so he knows what he has been told. And, trusting the speaker,
he comes to believe the very thing he has been told—that is, what the
speaker said and believed. Thus what Elwood believes, at the beginning
of the episode, is just what Alphonse believes, by the end of the episode.

Some philosophers, including me, find it useful to think of this sort
of episode in terms of the concept of *content*. Some states (beliefs,
paradigmatically) and events (assertions, paradigmatically) are con-
tentful; they can be assigned truth-conditions, or some other sort of
success conditions as their contents, and be evaluated as true or false,
accurate or inaccurate, successful or unsuccessful in virtue of whether
those conditions are met. Locutions such as "what Elwood believes"
and "what Elwood told Alphonse" identify such contents. Contents are
common to different kinds of meaningful states. A picture can show
that Sacramento is beautiful; a person can believe that it is, a traveler
can wonder whether it is, and hope that it is, a travel-agent can say

that it is, and so forth.

Contents seem to be rather amazing things, for several reasons:

- They are elements that are somehow in common to different properties, of different kinds of things— properties like showing that Sacramento is beautiful (a property of a picture), and believing or hoping that Sacramento is beautiful (quite different properties of people).

- They typically involve objects that are not a part of the things the properties of which they are used to characterize. Sacramento, for example, isn't part of the picture which has the property of showing Sacramento to be beautiful, nor is it part of Elwood or Alphonse, who have the property of believing it to be beautiful.

- In spite of this, contents are used to characterize states that we think of being local to, or even inside of, individuals. We think of Elwoods's belief as in his head, his utterance as an act of his. Alphone's hearing, understanding, and the belief he acquires are similarly involved with his head. What can Sacramento, a distant city, have to do with these internal states and events?

- We think of contents as having logical relations among themselves, which we use in explanations and inferences about these mental states that somehow incorporate them. Thus in the paragraph with which we began, we explain Elwood's saying something by reference to a belief with the same content. We would expect him to deny the inconsistent content that Sacramento was unattractive, and to believe the entailed content that some city was attractive.

The Philosophy of Content is a subject that cuts across the philosophy of language, the philosophy of mind, logic, and other parts of philosophy. In the twentieth century the philosophy of content has been an important part of analytical philosophy, although not until recently under that name. This stream of thinking began with currents from central Europe, especially Meinong and Frege. Russell was influenced by both of them, and issues we now can see as very much about content were in at the birth of analytical philosophy.

Much of the important work on the theory of content comes from philosophers who are more or less skeptical: Stich and Schiffer, for example. Others important figures have been, in various ways and in line with various paradigms, enthusiastic: Fodor, Lewis, and Searle for example.[1] I believe that one factor that has led to skepticism about content, and interfered in other ways with a properly positive theory of content, is the idea that for content to make sense, there must be a

---

[1]See the items listed in the bibliography for these authors.

certain kind of content, that I call the Semantic Grail of the philosophy of content. The Semantic Grail is just the content that examples like the one above lead us to expect: a single content that:

(A.) Is the content of Elwood's belief, his utterance, Alphonse's understanding, and Alphonse's belief.

(B.) By being the content of each of these mental states and linguistic acts, explains its connections with the causes and effects of that state or act.

In this paper, I claim there is no such content; there is rather a structure of related contents. One content fills role A. A family of other contents fills role B. Instead of a single grail, we have sort of a semantic tea service.

4.2   Referential Content and Cognitive Content

The natural place to begin our search is with report of Elwood's utterance.

(1) Elwood said that Sacramento is a beautiful city.

There is considerable (if far from universal) agreement among philosophers of language that this report says that the content of Elwood's utterance is a proposition about Sacramento. The proposition might be modeled as a set of possible worlds in all of which Sacramento was beautiful; it wouldn't be required that it be called "Sacramento" in all of the worlds, or be the capital of California, or be a town to which Elwood referred. Nothing would be required of the city except whatever is required to make it the very same city we in our world call "Sacramento, California".

One can also model this proposition simply as as a pair consisting of the city, Sacramento, and the property of being beautiful. We have what David Kaplan calls a "singular proposition"; that is, a proposition individuated by an individual and something asserted of itKaplan (1989).

I call this the "referential content" of the statement, because it incorporates the object referred to, and I call the thesis that such referential contents are what is said, "referentialism".

Referentialism and its singular propositions strike some people as rather mysterious. How can we express a proposition with a city as a constituent? Propositions should be made of things with which we can have a more intimate mental connection than we can with than cities. But there is a simple and non- mysterious interpretation of referentialism. We can think of the singular proposition as giving the truth-conditions of an utterance given the facts about reference. Given

that Elwood was referring to Sacramento, California with his use of the name "Sacramento", what *else* has to be the case for his utterance to be true? It has to be a beautiful city. Contents characterize an utterance by what the world must be like for it to be true. More or less can be taken as given; the proposition assigned as content gets at what else the world must be like. The fact that "what is said" is typically referential content, simply reflects our typical interest in certain kinds of facts about utterances: what the world has to be like for them to be true, given the language, the meaning and the reference of the words in them.

Here is our first candidate for the Semantic Grail of Content, then, the singular proposition that Sacramento is a beautiful city. The referentialist will think that this proposition is what Elwood believed, what Elwood said, what Alphonse took himself to have been told, and what Alphonse came to believe. This is the content that fills role A. Whether or not referential content is the Semantic Grail, it is at least an important part of the tea service.

Referential content does not, however, handle the explanatory burden that we were looking for with B. Elwood might have made many different statements that would have expressed this singular proposition, but that would not have been cognitively equivalent; that is, they would not be motivated by the same beliefs, and would not lead to the same beliefs on the part of a credulous speaker. If Elwood had been in Sacramento and said,

(2) This city is beautiful

he would have expressed the proposition that Sacramento is a beautiful city — at least according to Kaplan's classic account of the content of statements using indexicals and demonstratives Kaplan (1989). But notice that he might have said this before he realized that he was in Sacramento. If so, there will clearly be an important change in his belief states when he realizes he is in Sacramento. The earlier belief state would be expressed by (2). The belief state he is in after the realization, he could express with,

(3) Sacramento is beautiful.

Consider these pairs:

(B2) The belief Elwood acquires as he drives into Sacramento, not knowing where he is;

(U2) Elwood's utterance, "This city is beautiful".

(B3) The belief Elwood acquires when he realizes the city into which he has driven is Sacramento

(U3) Elwood's utterance, "Sacramento is beautiful".

The utterance (U2) is the natural expression of the belief (B2); the utterance (U3) is the expression of the belief (B3). But we cannot distinguish between the content of (U2) and (U3) on the basis of their standard referential semantics, for they are assigned the same singular proposition.

The hypothesis that the singular proposition expressed by Elwood is the one we are searching for does not explain the explanatory links between the steps. There are a number of states Elwood could have been in, that would have constituted his believing the singular proposition in question. There are a number of things he could have said, that would have constituted saying this. The singular proposition doesn't distinguish between these different beliefs and different utterances, and so doesn't really explain the links between them. This is, of course, just a version of Frege's problem with modes of presentation, a problem now into its third century (Frege, 1960).

4.3 Reflexive Content

The underlying problem with referential contents is that quite different conditions are required of a city to be the reference of "this city" in (U2), on the one hand, and to be the reference of "Sacramento" in (U3), on the other. To be the reference of the first, a city must be the one Elwood is in and is demonstrating when he uses the words "this city". To be the reference of the second, a city needs to be the reference of Elwood's use of "Sacramento". Elwood uses the term "Sacramento" to refer to the capital of California in a variety of situations; its reference does not depend on his being in or demonstrating Sacramento, even when he is in a position to do so. The knowledge or beliefs that motivate the use of "this city" are then different than the ones that motivate the use of "Sacramento". As Elwood enters the city, he acquires a belief about a certain city that it is beautiful. The city he is acquiring the belief about is the one he perceives. All he needs to know, in order to refer to the city he perceives with the words "this city", are the rules for the use of demonstratives in English. Even though Elwood is lost, he can manage this. But to express the thought that the city one is perceiving is beautiful by saying "Sacramento is beautiful," one needs to know or at least believe that the city one is perceiving is called "Sacramento".

It seems then that in order to get at the more fine-grained content of Elwood's belief that he conveys in his utterances, we need a way of looking at content that does not incorporate the referents of the names, but rather the conditions for being the referents of the names.

I provide such a level of content for utterances in my book *Reference and Reflexivity* Perry (2002). I call it "reflexive content". The idea is very simple; one looks at the truth conditions of an utterance without fixing the referent. Suppose I say,

(4) Today is cold

on July 1, 2000. Taking "Today" to be a term that refers to the day it is used, and taking into account when I used it, it seems I have expressed the singular proposition that July 1, 2000 is cold. What additional has to be the case for my utterance to be true, given that it is in English and uttered on July 1, 2000? That July 1, 2000 be cold.

But suppose we don't take the contextual fact, about when (4) was uttered, into account? Then it seems like we can say the following:

(5) (4U) is true iff the day on which (4U) is uttered is cold.

This gives us as a truth-condition of (4U) a proposition about (4U) itself:

(6) That the day on which (4) is uttered is cold

(6) is what I call the reflexive content of (4). "Reflexive" means simply that the truth-conditions of (4) are given in terms of conditions on (4) itself. The reflexive content corresponds to what someone understands who hears an utterance like (4) without knowing on what day it occurs. Also, since we take utterances to be intentional acts of writing as well as speaking, we can think of someone who finds (4) written in a diary, without any indication of when it was written. Such a person can be said to understand the utterance, in that they understand the language in which it was written, and know the conditions under which it would be true. But they don't really know what proposition was expressed. Notice that someone who heard Elwood make this statement, but had no idea what day it was, could nevertheless verify whether it was true or not. We have ways of finding out if it is cold on a given day, namely, stepping outside and seeing, or looking at a thermometer through the window.

This point about days and temperatures is similar to the one made above about cities and being beautiful. Elwood could tell that Sacramento was beautiful, just by looking around him, even if he had no idea what city it was, or thought it was Stockton or Chico (see Perry (2000). We can also imagine cases that go the other way around. Elwood might know that Sacramento beautiful, because he read this in an authoritative travel guide that he paid good money for. And he might be in Sacramento. But he might not know that the city he was in was beautiful, for he might not have gotten to the parts of Sacramento that

make it beautiful. Constructing a similar case about the weather is left as an exercise for the reader.

If we look back at (U2) and (U3), we see that their reflexive contents are quite different:

Reflexive Content of (U2): The city the speaker of (U2) demonstrates is beautiful.

Reflexive Content of (U3): The referent of the speaker's use of "Sacramento" in (U3) is beautiful.

These differences correspond to the differences we saw in the cognitive content of (U2) and (U3). Elwood, the speaker of (U2), needs to be able to demonstrate the city of which he speaks, but does not need to know its name. With (U3) Elwood needs to know the name of the city he asserts to be beautiful, but does not need to be able to demonstrate it. The reflexive content is not what Elwood said in either case. The subject matter of his utterance was a city, Sacramento, not the utterance itself, in both cases. If we reconstruct the implicit plans the speaker might have, the role the reflexive content plays becomes clearer. Suppose Elwood merely wants his companion to look up from the map for a moment and notice the pretty city they have entered. He wants his companion to think something like, "if I look out the window, I'll see a beautiful city". He plans as follows:

> I will utter "This city is beautiful," producing a token — a disturbance of the air waves — that will impinge on my companion's ears. My companion will perceive the disturbance as a token of an English sentence, the meaning of which he knows. At this point, he will think "The city the speaker of this token demonstrates is beautiful". It will be obvious to him that I am the speaker. He will look at me to see if I am demonstrating the city we are in, or perhaps pointing to one on the map, or perhaps to a sign, or whatever. Once he sees I am merely directing my eyes out the window of the car to the city we are in, he will realize, "If I look out the window, I'll see a beautiful city". And so, if he has any desire to see a beautiful city, he will take his eyes off the map for a second and look outside.

I have, of course, represented as a conscious plan that which would be quite below the level of consciousness in most cases. The premises reflect the steps that would be involved in the bit of know-how exhibited by speaker were they made explicit. In fact, most adults are pretty good at producing utterances that fit plans suited to the context they are in. Those who are not skilled, or who don't bother to exercise their skill, can be irritating. Such people expect you to see where they are pointing when you can't see them (perhaps they are in the back seat

of a car you are driving), to know who they are when they call on the phone, without telling you their name, to know which person has just popped into their mind, even though they refer to them with a "he" or "she", and so forth.

Uttering (3) would not have worked to get the result Elwood wanted. The companion might not realize that Sacramento is the town around them. He might just agree, saying, "Yes, I'm sure it is; I can't wait until we get there".

On the other hand, Elwood might plan to use (3) as a way of transmitting the information that they were in Sacramento. Suppose it was not common knowledge between Elwood and his passenger that Sacramento was beautiful, and Elwood had in fact expressed some skepticism on the point earlier. Elwood figures that when he utters (3) his companion will wonder what motivated the change of mind, and look at Elwood; seeing that Elwood was looking out the window and appreciating the town they were in, he will figure out the simplest explanation, that they are in Sacramento and Elwood has revised his opinion based on seeing it.

Our descriptions of the contents of utterances focus on their referential contents, rather than their reflexive contents. This is natural, since our conversational goal is usually to transmit information about the things to which we refer, not about the utterances we make. Elwood doesn't want the companion to remember that a particular utterance was made about a beautiful city, but that a particular city was beautiful. The belief about the utterance was a stepping-stone to this belief, which can quickly be forgotten. Nevertheless, we clearly are adept at planning our utterances in ways that exploit their reflexive contents, and provide appropriate stepping stones for the hearer to get to the belief we are aiming for her to have.

Does the reflexive content of an utterance then provide what we need for aspect B of the Semantic Grail? It cannot, for the reflexive content of Elwood's utterance is not the same as the reflexive content of the belief that motivated it and that it expressed, nor of the belief that Alphonse acquired.

The belief was present in Elwood's mind before he made the utterance, and would remain there even if he had decided not to say anything. If a content is to be the reflexive content of the belief, it must be reflexive, that is, place a truth condition on the belief itself, not on an utterance, particularly one that might not exist yet. The reflexive content of the utterance is another piece of our tea service, but not everything we need.

4.4 Doxastic Content

We can make good sense of the idea of the reflexive content of a belief: the proposition that gets at the conditions truth of the belief puts on the belief itself. To do so, however, we need a model of what beliefs are like. I'll assume simple beliefs like those expressed by (U2) and (U3) consist of ideas being connected together in the mind. Ideas of properties and relations I'll just call *ideas*; ideas of individuals I'll call *notions*. I'll further assume that notions can be attached to or detached from perceptions. The notion of Sacramento involved in the belief that led to (U2) was attached to the speaker's perception of Sacramento. Information from the perception was influencing which ideas were associated to the notion. If we imagine (U3) being uttered by someone far away from Sacramento, perhaps planning a trip, the notion would be detached. The notion is associated with various ideas gleaned from reading about Sacramento and perhaps from memories of previous visits. But no present perception of Sacramento is feeding information into the complex of notion and associated ideas.

Let $\mathbf{n}_{B2}$ be the notion that is part of the belief B2 and is attached to Elwood's perception. It is associated with the idea of being beautiful. Stating the truth-condition of B2 in terms of its own constituent, the notion $\mathbf{n}_{B2}$, we have:

Reflexive Content of (B2): That the city that the perception attached to $\mathbf{n}_{B2}$ is of, is beautiful.

Let $\mathbf{n}_{B3}$ be the notion that is part of the belief (B3) and is associated with the idea of being named "Sacramento". This notion is also associated with the idea of being beautiful. Stating the truth- condition of B3 in terms of its own constituent, the notion $\mathbf{n}_{B3}$, we have:

Reflexive Content of (B3): That the city that the notion attached to $\mathbf{n}_{B3}$ is of, is beautiful.

Note that $\mathbf{n}_{B2}$ and $\mathbf{n}_{B3}$ might be the same notion, or might be different notions that are linked so information passes between them, or might be different and unlinked. They will be the same or linked if the speaker realizes that the city he is demonstrating is Sacramento. The speaker will know how to refer to a city his notion of which is attached to a perception, for that is just my theoretical jargon for knowing how to refer to the city he sees. He will know how to refer to a city his notion of which is associated with the name "Sacramento," for that just requires knowing that you can refer to a thing with its name.

A reconstruction of the process that leads from belief to utterance thus needs to bring in auxiliary beliefs. In the case of (U3), the speaker

believes not only that a certain city is beautiful, but also that the same city is named "Sacramento". In the case of (U2), he believes not only that a certain city is beautiful, but also that the city he will refer to with "this city" is that very one. When the auxiliary beliefs are wrong, the speaker does not say what he intended to.

In the case of (U2), if the speaker is merely perceptually picking up the information that Sacramento is beautiful, and then passing it on demonstratively, it's hard to imagine him not saying what he intends to say. Another scenario, however, is that he believes Sacramento to be beautiful quite independently of his current perceptions. He has believed this for years, since he first visited Sacramento. He intends to express *this* belief with (U2). He thinks he can refer to the city he takes to be beautiful by saying, "this city," for he thinks he is in Sacramento. But in fact he is in Chico. When he says "This city is beautiful," he does not say what he planned to say. Or, more accurately, his plan involved saying one thing by saying another; the first part succeeded but not the second. Elwood thought that by expressing the proposition, about the city he was looking at, that it was beautiful, he could express the proposition that Sacramento was beautiful. He successfully carried out the first part of the plan; he expressed the proposition about the city he is looking at, that it was beautiful. But he doesn't get to the goal. He doesn't thereby express the proposition that Sacramento is beautiful.

The motivating belief, the one which one intends to express, and the motivated utterance do not then have the same reflexive content. It will be handy, however, to have the concept of the *doxastic content of an utterance*. Where b is a belief that u is intended to express, the reflexive content of b is the doxastic content of u.

The explanations that are implicit in the opening story about Elwood and Alphonse are based on a common situation. We suppose that Elwood's utterance was intended to express his belief. We assume that there is a correspondence between notions in the belief and terms in the utterance (See Crimmins and Perry (1989)). For an utterance to express a belief, each notion must be of the same thing as the corresponding term refers to, and the predicate must predicate the property the corresponding idea is of. If these conditions are met, the reflexive content of the utterance and the doxastic content of the utterance will be referentially equivalent. That is, the referential content that we get by adding the referential facts about the utterance to what is given, is the same as the referential content that we get by adding the facts about what things the notions in the motivating belief are about. What is believed is what is said.

The hearer will typically not rest content with grasping the reflexive

content of the utterance, but will make an identification between the object spoken about and some object of which he has his own notion, if he can. So in the case of (U2) the hearer takes the city referred to by the speaker's use of "this city" to refer to the city that the hearer sees, since it is the same one that the speaker is demonstrating via eye gaze through the car window. The hearer will then acquire a belief whose reflexive content will be referentially equivalent with the reflexive content of the utterance. If the auxiliary semantic assumptions are wrong, this will not be so.

4.5 A story of four contents

There is no Semantic Grail of Content, no Fregean sense that characterizes the cognitive content of each of the parties to an episode of communication, and is also what is believed by the speaker, understood by the hearer, and expressed by the utterance. In a successful act of communication, there will be a single proposition expressed by the utterance, and believed by both participants, the referential proposition. But it won't get at the cognitive content.

The reflexive contents of the two beliefs and the utterance are more closely connected with cognitive content. But the situation is not as simple as having a single reflexive content for all three. The reflexive contents have architectural and explanatory connections. The sincere speaker plans to utter something that will express the proposition he believes. This means that the truth-conditions of the utterance, given the facts that determine the meaning and reference of the words in it, should be the same as the truth-conditions of the motivating belief, given the facts that determine what the ideas and notions in it are about. The are a number of ways to do this. There are ways of expressing exactly what is on your mind, without taking a chance of depending on any further facts one might get wrong. But the risk-aversive strategies may not advance one's conversational goals.

Let's go back to Elwood and his passenger. Elwood has trouble keeping the names of Stockton, Sacramento and other central valley cities straight. Thinking of Stockton, but fearful of getting the name wrong, Elwood could simply say, "It's beautiful". The "It" inherets its reference from the thought, so the utterance has the same referential content as the motivating belief. And this is what we sometimes do, when we can't remember the name of the object about which we are thinking—or perhaps we know, but we want the hearer to guess what we are talking about. This cautious strategy would not be very helpful to the passenger; he will know that Elwood is talking about the thing he is

thinking about, but won't have any idea which object that is.

Think now of Elwood in the car, seeing Sacramento, but not knowing which city he sees. Elwood knows that in English one can use "this" and "that" to refer to the object to which one is attending. So he says, "That city is beautiful" to his passenger which correctly expresses his thought. Not too much can go wrong here. That is because the way he is referring is directly related to his perceptual mode of belief. When things go right, a person's linguistic competence, plus other relevant auxiliary beliefs, should guarantee the sameness of the referential contents of motivating belief and utterance. The sameness of referential contents of utterance and acquired belief should similarly be guaranteed by linguistic competence and auxiliary beliefs.

Although there is no single content to serve as our Semantic Grail, there is a structure of contents in a communicative interchange whose contents are systematically connected when things work correctly. If there are no misunderstandings, the reflexive content of the motivating belief plus auxiliary beliefs will explain the reflexive content of the utterance, which will in turn, together with auxiliary beliefs of the hearer, explain the reflexive content of the acquired belief. And both beliefs and the utterance will have the same referential content.

References

Crimmins, Mark and John Perry. 1989. The prince and the phone booth. *Journal of Philosophy* pages 685–711. Reprinted in Perry (2000).

Fodor, Jerry A. 1990. *A theory of content and other essays*. Cambridge, Mass.: MIT Press.

Frege, Gottlob. 1960. On sense and reference. In M. B. Peter Geach, ed., *Translations from the Philosophical Writings of Gottlob Frege*, pages 56–78. Oxford: Basil Blackwell. Translated by Max Black.

Kaplan, David. 1989. Demonstratives. In H. W. Joesph Almog, John Perry, ed., *Themes from Kaplan*, pages 481–563. New York: Oxford. Manuscript circulated in 1977.

Lewis, David K. 1979. Attitudes *de dicto* and *de se*. *Philosophical Review* 88:513–43.

Meinong, A. 1960. The theory of objects. In R. Chisholm, ed., *Realism and the Background of Phenomenology*, pages 76–117. Glencoe: The Free Press.

Perry, John. 2000. *The Problem of the essential indexical, enlarged edition*. Stanford: CSLI Publications.

Perry, John. 2002. *Reference and reflexivity*. Stanford: CSLI Publications.

Russell, Bertrand. 1973. Meinong's theory of complexes and assumptions. In D. Lackey, ed., *Essays in Analysis*, pages 21–76. New York: George Braziller.

Schiffer, Stephen R. 1987. *Remnants of meaning*. Cambridge, Mass.: MIT Press.

Searle, John R. 1983. *Intentionality, an essay in the philosophy of mind*. New York: Cambridge University Press.

Stich, Stephen P. 1983. *From folk psychology to cognitive science: The Case Against Belief*. Cambridge, Mass.: MIT Press.

March 22, 2006

5

On a Proposal of Strawson Concerning Context vs. 'What Is Said'

Varol Akman

5.1 Introduction

In[1] human communication using natural language, there is potential for a certain intricacy regarding the communicative mode and 'what is said' (Ziff, 1972). The following anecdote comes from Johnson-Laird (1990, p. 7):

> Once upon a time Stalin read out in public a telegram from Trotsky: 'You were right and I was wrong. You are the true heir of Lenin. I should apologize. Trotsky.' According to Leo Rosten, a Jewish tailor then stepped from the crowd and explained to Stalin how he ought to have read the message:
> You were right and I was *wrong*? *You* are the true heir of Lenin? *I* should apologize???!!

While one appreciates the crucial role of intonation in this story, most of us also realize that it is the historical background through which the intended meaning is contextually determined in this case. Thus, Stalin's rendering of the individual words or phrases do make sense but it is the 'deconstructive' reading of the tailor that goes to

show the determinate meaning of Trotsky's message.

But, how does one *really* purport to know the intended meaning of a given message? Few would deny that in the process of construing meaning, one is caught in the act of contextualizing—placing things in context.[2] In studying this act in any detail, it is unavoidable to notice the interaction of authorial intentions and context. And it turns out that in his most recent volume of essays, Strawson has considerable things to say on this very question.

Strawson's book is entitled *Entity and Identity*, and the essays which treat the afore-mentioned question at some length appear as Chapters 11 and 12 (Strawson, 1997a,b). In these essays,[3] Strawson advances a particularly attractive threefold distinction regarding how context bears on the meaning of 'what is said' when a sentence is uttered (Ziff, 1972). In his view, three senses (sense-A-meaning, sense-B-meaning, and sense-C-meaning) capture increasingly more intricate and progressively richer aspects of what is said. But Strawson cautions that his proposed scheme may still be simplistic, since the situation may be more complicated than the scheme suggests, and raises various points to make it more adequate.

In this paper, we'll (i) review the original scheme of Strawson and summarize his improvements to his own scheme, and (ii) add our own improvements to make it even more thoroughgoing. Overall, we'll defend the versatility of Strawson's framework. On the other hand, we'll also show that unless it is elaborated with several considerations (mostly based on a viewpoint regarding context as a social construct (Akman, 2000) and contextualizing as a form of social action) it cannot function as a realistic initiative towards building common sense models of how intended meaning is achieved.[4]

---

[2]Modern literary theory distinguishes between an author's intended meaning and whatever *significances* a reader finds in the text. Not all patterns and relationships found by the reader in a text can be attributed to authorial intentions . The producer of a text, Eco (1984, p. 7) claims, "has to foresee a model of the possible reader [...] supposedly able to deal interpretatively with the expressions in the same way as the author deals generatively with them." This possible reader Eco calls the *model reader*. In order to make his text communicative, the author has to make sure that the totality of 'codes' upon which his work is built is the same as that shared by the model reader.

[3]The original essays were published considerably earlier. Thus, Chapter 11, "Austin and 'Locutionary Meaning'," first appeared in Isiah Berlin et al., eds., *Essays on J. L. Austin*, Oxford University Press (1973). A partial translation of Chapter 12, "Meaning and Context," appeared in *Langages* 17 (1970), with the title "Phrase et acte de parole."

[4]An explanation regarding the motivations of the two essays is in order. Austin (1976) famously distinguished between the *meaning* and *force* of an utterance. He

5.2   Strawson's Scheme

Strawson tackles the riddle of how context influences intended meaning by first proposing a simple question and an economic answer. He then attends to the complications which seem not to be easily resolvable by the latter.

Assume that a certain sentence S of a language L (e.g. English) was seriously uttered on some occasion. (N.B. The adverb "seriously" plays a crucial role, as we'll later see.) Assume further that X, the hearer, possesses only that much information, i.e. X knows that S was uttered but knows nothing about the identity of Y, the speaker, or the nature or date of the occasion. (In various places in the sequel, this restriction will be relaxed.) Let us grant X full mastery of the syntax and semantics of L; thus, X is assumed to have ideally complete knowledge of L (lexicon plus grammar). The question is as follows (Strawson, 1997a, p. 192):

> [I]s there any sense in which X can be said to know *the meaning of precisely what was said on the occasion in question*?

Strawson's proposed scheme to investigate this problem consists of erecting three progressively richer senses of meaning which he dubs sense-A-meaning, sense-B-meaning, and sense-C-meaning.

5.2.1   Sense-A-Meaning

Sense-A-meaning is *linguistic* meaning. Suppose S is free of ambiguity, or more realistically, X is informed which of the alternative readings of S is the right one, i.e. the one meant by Y. (It is beside the point, for the time being, *how* X could be told which of the possible lexical items or syntactic constructions Y actually had in mind in uttering S.) We then say that X knows the *sense-A-meaning* of 'what is said'.

An important characteristic of such meaning is that if he has access to it, then X can give a correct translation of S into another language L' (e.g. French), which X, once again, is assumed to know perfectly well. In other words, when sense-A-meaning is under consideration, X basically knows neither more nor less than what he needs to know in order to translate S into a sentence S' of L'.

Consider the following example (due to Strawson) as S: "The collapse of the bank took everyone by surprise." The designation of the word "bank" varies with different uses. But once the intended des-

---

associated the former with the 'locutionary' act performed in making the utterance, and the latter with the 'illocutionary' act. In his chapter on Austin, Strawson *uses* the threefold distinction to examine Austin's work; his standpoint is that what Austin means by locutionary meaning is not very clear. On the other hand, in "Meaning and Context" the threefold distinction *itself* is examined in detail. (Our remarks will generally bear on the contents of this essay.)

ignation is clarified, then the translation of S from L to L' proceeds smoothly. Also witness Perry's similar remarks (Perry, 1998, p. 2): "An ambiguous expression like 'bank' may designate one kind of thing when you say 'Where's a good bank?' while worried about finances, another when I use it, thinking about fishing. [...] Is the speaker holding a wad of money or a fishing pole?"

To summarize the preceding paragraphs,

sense-A-meaning $\approx$ S $\oplus$ A-knowledge $\oplus$ disambiguating knowledge,

where *A-knowledge* is the ideally complete knowledge of the lexicon and grammar of L. In this mock equation, the interpretations of "$\approx$" and "$\oplus$" are somewhat procedural; that is, the equation states that sense-A-meaning is obtained (approximated, if you will) by just understanding S in the light of A-knowledge and disambiguating knowledge (and with a propensity toward accurate translation of S into any other, equally rich language).

5.2.2   Sense-B-Meaning

Strawson's sense-B-meaning is *linguistic-cum-referential* meaning. X will learn the *sense-B-meaning* of S if he has access to the references of proper names or indexicals which may be contained in S.

An example might illustrate the difference between sense-A and sense-B meanings. If S is the sentence "He stood on his head since then," and if X is further told that this potentially ambiguous sentence has its natural reading where "his" is co-indexed with "he," then X can easily translate S to say, French. When X does that accurately, it would show that X understood the sense-A-meaning of S. Now suppose X has no idea who "he" stands for and which time point "then" denotes. This might not pose a problem for the translation. But if X additionally learns the reference of "he" (say, J. L. Austin) and "then" (say, New Year's Day, 1955) then X would know a richer meaning, the sense-B-meaning of S.[5]

In a style suggested by the earlier equation,

sense-B-meaning $\approx$ sense-A-meaning $\oplus$ B-knowledge,

where *B-knowledge* includes—in addition to A-knowledge—the knowledge of the reference of proper names and indexical expressions that

---

[5]One may object to the preceding analysis by noting that there are naturally occurring contexts in which the particular S of this example might have metaphorical meaning. Or at least, this is exactly what happens when one replaces S with a similar sentence "He stood on his own feet since then," meaning: he thought and acted independently since then. We agree and note that this is precisely the point of Strawson's imposition, viz. S is uttered seriously. More on this later.

might be occurring in S. Again, this mock equation can be interpreted as follows: sense-B-meaning is obtained by scrutinizing sense-A-meaning in the light of B-knowledge.

5.2.3 Sense-C-Meaning

Finally, Strawson offers sense-C-meaning as the *complete* meaning of a message. *Sense-C-meaning* is obtained by adding to sense-B-meaning the illocutionary force (à la Austin) of what was said, together with a complete grasp of how what was said is intended (by Y) to be understood (by X). Thus,

$$\text{sense-C-meaning} \approx \text{sense-B-meaning} \oplus \text{C-knowledge},$$

where *C-knowledge* consists of—in addition to B-knowledge—the illocutionary force of S plus the true intent of Y. For instance, if S is the sentence "Don't sign that contract yet," then X needs to know whether this was issued as a request, a command, a piece of advice, or what have you. This is the dimension of meaning Austin captured with the phrase 'illocutionary force'.[6]

There is a related but distinct notion: it may be that Y intends to be taken to be implying by S something which does not ensue from S's sense-B-meaning alone. Assume that both X and Y know (and know each other to know) that their mutual friend Z declined an honor conferred upon him by a church. When Y says "It is the sign of a feeble mind to turn down a gift from God," the meaning of what he said would not be fully understood by X if X fails to recognize that Z is being labeled as the decrepit one by Y. Grice (1989) was in some sense the first to produce an elucidation of how a speaker can communicate more than what his words explicitly say. Since Strawson does cite Grice, it is safe to predict that he has in mind the same kind of systematic Gricean principles underlying pragmatic 'implication'.

5.2.4 An Inequality

With the preceding three equations at hand, we can write the mock inequality

---

[6]Strawson understands illocutionary force as having to do simply with what the speaker means. However, illocutionary force has to do with what is conventionally constituted by the locutionary act being performed in context. Witness the following caveat of Austin (1976, pp. 116–117): "I cannot be said to have warned an audience unless it hears what I say and takes what I say in a certain sense. [. . .] So the performance of an illocutionary act involves the securing of *uptake*." In a nutshell, then, one decides on what interpretation to accept by examining uptake—the (conversational) process through which lines of reasoning are developed/modified (Gumperz, 1997).

$$\text{sense-A-meaning} \prec \text{sense-B-meaning} \prec \text{sense-C-meaning},$$

where progressively richer senses of meaning are obtained by moving from left to right in the inequality. Since X employs (in proceeding from S to sense-A, sense-A to sense-B, and sense-B to sense-C) A-knowledge, B-knowledge, and C-knowledge, respectively, the progression in meaning will in general be *additive.* However, sometimes the move from one sense to another is really no move at all. A fitting example comes from mathematics: let S be a sentence expressing a proposition of arithmetic, e.g. "There is always a prime number greater than a given natural number." In this case, the move from sense-A to sense-B is no move at all because the statement S expresses an analytic truth.

How about C-knowledge? Can its contribution also be null sometimes?[7] The answer is not in the affirmative, despite what Strawson thinks.

To see this, take an explicitly performative statement such as "I order you to drop that gun." Together with Strawson, we may, at first, be inclined to accept that knowledge of the force of this S can be taken to belong to the sense-A-meaning. However, this is not really to follow Austin (1976). To give an example, if a mutinous private in the British army purported to order his sergeant to drop his gun and the cowardly sergeant did so, then a court martial would definitely rule that there was no order (or nothing with the force of an order), because a private cannot give an order to a sergeant. In other words, it is one thing for a type to be meant to be tokened in an act with a certain force and another thing for the token *actually* to realize an act with that force.

5.2.5   Leech's Scheme

Another threefold distinction due to Leech is worth indicating at this point. Leech states that specification of context has the effect of narrowing down the communicative possibilities of a message. He says that in particularizing meaning, context helps in the following ways (Leech, 1981, p. 67):

· (A) Context eliminates certain ambiguities or multiple meanings in the message (e.g. lets us know that *page* in a given instance means a boy attendant rather than a piece of paper).

· (B) Context indicates the referents of certain types of word we call deictic (*this*, *that*, *here*, *there*, *now*, *then*, etc.), and of other expressions of definite meaning such as *John*, *I*, *you*, *he*, *it*, *the man.*

---

[7]In which case the move from B to C might still be regarded as an addition, even if it is the minimal addition that *there is nothing to be added* to the B-meaning.

- (C) Context supplies information which the speaker/writer has omitted through ellipsis (e.g. we are able to appreciate that *Janet! Donkeys!* means something like 'Janet! Drive those donkeys away!' rather than 'Janet! Bring those donkeys here!', or any other of the indefinitely many theoretical possibilities).

Clearly, (A) states the so-called *disambiguating role* of context and immediately brings to mind Strawson's sense-A-meaning. Likewise, (B) is along the lines of Strawson's sense-B-meaning. Finally, although the singling out of ellipsis might at first sight seem way too specific, it is clear that Leech is talking in (C) about a particular way of how speaker's intention is to be inferred. His example has the same import as Strawson's sense-C-meaning, viz. the requirement that the reader must be aware of all that was intended by the speaker. "Janet! Donkeys!" is recurrently used by aunt Betsey Trotwood in *David Copperfield*; it is an order to her maid to carry out the routine task of driving donkeys off the grass.

5.3 Dependence on Context

Having defined the three senses of meaning, A-, B-, and C-, Strawson turns to the following question: what specific differences are there in the ways in which the meaning of 'what is said' depends on context in the three cases? In particular, in which cases and to what degree can this dependence be itself represented as governed by *linguistic rule* or *convention*?

Obviously, context bears on determination of sense-A-meaning in just those situations where S suffers from syntactic and/or lexical ambiguity. However, disambiguation of S by context at this level is not in general a matter of linguistic rule or convention. Rather, it is a matter of general relevance; see the earlier example of Perry regarding which meaning of "bank" might be more plausible. In the same vein, Leech (1981, p. 69) states that it is relevant to the interpretation of "Shall I put the sweater on?" to know whether sweaters heated by electric power are on the market. This shows, in a rather strong sense, that the study of interpretation-in-context is closely tied to the *encyclopedic knowledge* about the world.

Context bears on the determination of sense-B-meaning in all cases except those where B-knowledge adds nothing to A-knowledge. And surely there are some semantic rules of natural language moderating such contextual dependence. Here's what Perry says about indexicals (1997, pp. 597–598):

There is an intimate connection between the meanings of "I" and "the

person who utters this token", even if it falls short of synonymy. The second phrase does not *have* the meaning of "I", but it *gives* part of the meaning of "I". It supplies the condition of designation that English associates with "I". [. . .] Here are the conditions of designation for some familiar indexicals [. . .]:

• **I**: u [an utterance of "I"] designates x iff x is the speaker of u
• **you**: u [an utterance of "you"] designates y iff $\exists x(x$ is the speaker of u & x addresses y with $u)$
• **now**: u [an utterance of "now"] designates t iff $\exists x(x$ is the speaker of u & x directs u at t during part of $t)$
• **that** Φ: u [an utterance of "that Φ"] designates y iff $\exists x(x$ is the speaker of u & x directs u towards $y)$

It is noted, however, that B-knowledge is not wholly under the governance of language rules (cf. Perry's caveat above: ". . . part of the meaning. . ."). For instance, with the demonstrative "here" there arises the question of how large a region to consider: "It is always very hot here at this time of the day" ("here": in this room or in this town?).

Similarly, an utterance of "We must sell those HAL stocks now" would signify different time points when it is made by a portfolio manager sitting at his on-line terminal ("now": in a couple of seconds) and by an executive during a luncheon with his assistants ("now": this afternoon).

5.4   Amendments

Strawson enumerated several points at which his threefold distinction is too crude to provide for all the complexities of language use. Despite the title of this section, he did not always suggest these as amendments to his scheme; sometimes he was content with just jotting them down.

5.4.1   Semantic Creativity

According to the inequality given earlier, some sense-A-meaning is always included in the complete meaning of 'what is said'. This is due to the nature of construction of sense-C-meaning. However, isn't it unrealistic to suppose that all meanings of a particular word are listed priorly in X's ideal lexicon? Consider the interpretation of a morphologically complex word w. Word formation rules might constrain but do not fully determine the interpretation of w. To put it mildly, the linguistically specified meaning of w may and frequently does go beyond what is available from its compositional subparts (Chierchia and McConnell-Ginet, 1990, pp. 366–370).

For example, Aitchison (1997, pp. 16–17) remarks that newspapers can popularize new words such as *yomp* and *wimp*. Yomp (to march

with heavy equipment over difficult terrain) was a military term used frequently during the Falklands War. Wimp (an ineffectual person) originated in the U.S.; just remember your generic White House correspondent during the Gulf War: "President Bush has finally shaken off his wimp image."

Récanati uses the term *contextual sense construction* to refer to the general problem. He notes that sometimes the conventional sense of the subparts of a complex phrase and the way they are syntactically brought together is insufficient to evaluate the semantic value of the complex phrase. His examples are particularly forceful (Récanati, 1994, p. 343):

> Thus 'he finished the book' can mean that he finished reading the book, writing it, binding it, tearing it into pieces, burning it, and so forth [...]; 'finger cup' will mean either 'cup having the shape of a finger' or 'cup containing a finger of whisky' or 'cup which one holds with one finger', or whatever [...]; 'John's book' can mean 'the book that John owns, wrote, gave, received', or whatever [...]. In all such cases there is not a 'selection' from a *limited* range of *preexisting* interpretations for the complex phrase. Rather, an indefinite number of possible interpretations can be constructed in a *creative* manner. [our italics]

Strawson finds his scheme too simple when it comes to such matters of semantic creativity. A compromise can be made by allowing X's ideal dictionary be updated by adding the new (extended) meaning of a new word. However, he sees this as a sacrifice of his ground rules: when we do this, we make X's dictionary follow his understanding rather than his understanding obey his dictionary.

5.4.2 Seriousness

Let us return to a crucial proviso in the original formulation, i.e. that a certain sentence S of a language L be *seriously* uttered. This implies that an *ironical utterance* of S is regarded as non-serious. However, ironical utterances make up quite a large crowd and cannot be so easily dismissed as aberrations.

The essential problem posed by ironical utterances is that a declarative sentence uttered ironically may express an idea that contradicts the idea which it professes to express. Consider saying "Oh, you are always so tidy!" to a janitor and meaning that he has made a mess again. Or consider related variants such as *understatements*, e.g. saying "It was rather concise" and meaning that it (e.g. a televised speech by the president) was extremely terse.

As (Strawson, 1997b, p. 222) notes, in these cases "we cannot say that the C-meaning includes and adds to the B-meaning, but only that

the C-meaning *contradicts* the *apparent* B-meaning."

Figurative uses pose a similar problem. Harris (1996, p. 112) says:

> If I say "Miller pulls off these tricks with string and sealing wax, false bottoms and sleight of hand," the statement will not be taken as figurative if I'm referring to an amateur magician, though it will be if the context makes it clear that I am referring to J. H. Miller the critical theorist.

5.4.3 Reference

Reference has always been a grand issue in studies of contextualism in the philosophy of language, and it is only normal that Strawson notes that sometimes a given S admits different interpretations where in one interpretation a certain constituent of S (e.g. a definite description) has a referential use whereas in some other interpretation it doesn't.

Suppose we pick as S the sentence "The next parliamentary elections will resolve the matter." The descriptive phrase may be used to refer to a definite event (say, the elections scheduled to June 8, 2004) or S may be used with the intention of saying "Whensoever the parliamentary elections are carried out, the matter will be resolved."

5.4.4 Translation Proper

An Italian saying, "Traduttore, traditore" (The translator is a betrayer), hints at the potential problems one can encounter in acquiring sense-A-meaning.

In a landmark essay on translation, Jakobson (1992) distinguishes three ways of interpreting a verbal sign. Intralingual translation (rewording) interprets verbal signs by means of other signs of the same language. Interlingual translation (translation proper) interprets verbal signs by means of some other language. Finally, intersemiotic translation (transmutation) interprets verbal signs by means of signs of nonverbal sign systems.

In order to demonstrate the difficulty of translation proper, he gives an example from Russian (Jakobson, 1992, p. 148):

> In order to translate accurately the English sentence "I hired a worker," a Russian needs supplementary information, whether this action was completed or not and whether the worker was a man or woman, because he must make his choice between a verb of completive or noncompletive aspect [. . .] and between a masculine and feminine noun [. . .]. If I ask the utterer of the English sentence whether the worker was male or female, my question may be judged irrelevant or indiscreet, whereas in the Russian version of this sentence an answer to this question is obligatory. On the other hand, whatever the choice of Russian grammatical forms to translate the quoted English message, the translation

will give no answer to the question of whether I "hired" or "have hired" the worker, or whether he/she was an indefinite or definite worker ("a" or "the").

5.5   Further Points

The following are not so much weaknesses of Strawson's scheme as possible avenues of research for streamlining it.

5.5.1   Radical Interpretation and Presemantic Uses

Regarding sense-A-meaning, the following singularity needs to be noticed: if his A-knowledge is null then X cannot even set himself to the study the question properly. This remark should not be taken as an avowal of the impossibility of radical interpretation. When X is a *radical interpreter* who must interpret L from scratch, he must do so in the absence of any antecedent understanding of L, and only using evidence which is plausibly available to him (Davidson, 1984).

That this is difficult, on the other hand, is something even Davidson himself accepts to a large extent (Kent, 1993): "It would beg the question, in trying to study the nature of interpretation, to assume that you know in advance what a person's intentions, beliefs, and desires are. [...] There is no master key or framework theory that you can have prior to a communicative interaction or situation."

Sometimes context is used to figure out which language is being spoken. Consider a well-known example due to Perry (2000, p. 314):

Ich! (said by several teenagers at camp in response to the question, "Who would like some sauerkraut?")

Perry says that knowing that this took place in a German rather than an American camp might help one to see that it was made by eager German teenagers rather than American teenagers repelled by the very idea. In this case, context (or rather its *presemantic* use) is pertinent to figuring out which language is being used.

5.5.2   Contextual Domains and Subjective Adjectives

A discussion given in (Récanati, 1998) refers to the fact that natural language quantifiers often seem implicitly restricted. When S is the sentence "The president shook hands with everyone," X is inclined to think that "everyone" must range over the domain of people who attended the press conference or the reception or the fund-raising dinner or whatever—not everyone in the whole world. Along similar lines, when Y utters "Most beggars attended the bash" he is likely to allude to a particular group of beggars (say, those in his neighborhood); it is from this group that many joined the festivities.

In dealing with *subjective* (*relative*) adjectives such as "large," the context contributes to meaning in a decisive way. Consider this (Chierchia and McConnell-Ginet, 1990, p. 374): "Lee built a large snowman." If Lee is a toddler playing in the backyard of his house, the snowman is probably at most as big as Lee himself. On the other hand, if Lee is a teenager competing in a snow carnival, the snowman is probably much bigger than Lee.

One way of dealing with the context-dependent nature of relative adjectives is to assume that the context provides us with a set of comparison classes. Still, with sentences like "A large tadpole is not a large animal" the problem remains unresolved; in the same context different comparison classes are needed for the first and second occurrences of the adjective.

5.5.3   Context Renewal

Consider an on-going conversation between X and Y. Y utters S, X in return utters S', Y in return utters S'', and so on and so forth. In order to understand say, S'', X would need to use the previous discourse, or the meaning of 'what was said earlier'.

That an interactional context is continually being developed with each successive utterance is an observation Heritage (1984) has made in his work on ethnomethodology. According to him, utterances and the social actions they embody are treated as doubly contextual. First, utterances and actions are *context-shaped*. This means that their contributions cannot be adequately appreciated unless the context in which they operate is taken into account. Second, utterances and actions are *context-renewing*. Every utterance will form the subsequent context for some following action in a sequence; it will thus contribute to the contextual framework which lets one understand the next action. Additionally, each action will function to renew context, where renewal is understood as one or more of the processes of maintaining, adjusting, altering, and so on.

In the remainder of this section we look at contributions similar in nature to Heritage's. Our general point is that at the level of sense-C-meaning Strawson's scheme would benefit from enhancements of sociocultural nature.[8]

5.5.4   Communicative Competence

Gumperz (1997, pp. 40–41) regards communicative competence as "the knowledge of linguistic and related communicative conventions that speakers must have to initiate and sustain conversational involvement."

---

[8]See (Fetzer and Akman, 2002) for recent work on social aspects of context.

This requires knowledge of social and cultural rules of a language (in addition to a knowledge of grammatical) and preferably addresses the competences of actual speakers, not an idealized standard.

Gumperz (1992) introduces what is known as a *contextualization cue*. He confirms that a given aspect of linguistic behavior (e.g. lexical, prosodic, phonological, etc.) can function as a cue, indicating those aspects of context which are to be taken into account to interpret what is said by a speaker. Contextualization cues hint at relevant aspects of the social context (via particular codes, styles, and dialects), thus enabling participants in a discourse to reason about their respective communicative intentions and purposes.

He also notes that because of its cultural base, the meaning of a conversation is frequently different for different participants if they are not members of the same speech community. Gumperz (1993) offers a case study of how differences in the use of contextualization cues between a native speaker of English and a non-native yet fluent speaker of English cause a serious breakdown in communication.

As another example of a cross-cultural communicative event, Saville-Troike (1989, pp. 131-132) observed the following exchange in a kindergarten on a reservation:

> A Navajo man opened the door to the classroom and stood silently, looking at the floor. The Anglo-American teacher said 'Good morning' and waited expectantly but the man did not respond. The teacher then said 'My name is Mrs. Jones,' and again waited for a response. There was none.

The whole exchange is more enlightening but this brief excerpt will serve to illustrate our point. The man's silence is appropriate from a Navajo perspective; it shows respect. What is more, a religious Navajo taboo prohibits individuals from saying their own name. Mrs. Jones's expectation is also reasonable from an Anglo-American perspective; the man must have returned her greeting, identified himself, and stated his reason for being there. It turns out that he was there to take his son, Billy, and that Billy is more accustomed to the Anglo-American ways than his stoic father. As he walks towards his father he waves at Mrs. Jones and says 'Bye-bye' to which she responds similarly.

5.6   Conclusion

The originator of a message usually assumes quite a bit of background knowledge on the part of an addressee (Leech, 1981, p. 66). The task of the addressee is to narrow down the list of meanings available to him and attain the intended meaning. Originally, the message may be re-

plete with several potential meanings. By 'enveloping' it in increasingly narrower contexts, the number of meanings is reduced. Eventually, it is hoped that just one meaning is isolated as *the* meaning of the message.[9]

This paper argued that there is a certain persuasive approach to studying the feasibility of this problem, first spelled out in "Austin and 'locutionary meaning' " and later taken up in detail in "Meaning and context," two early papers by Strawson. The approach is both simple and elegant, and we believe that future studies to formalize context (Akman and Surav, 1996, 1997) might profit from its formulaic nature.

Acknowledgments

An earlier version of this chapter was presented in *CONTEXT'99* (Akman and Alpaslan, 1999) and benefited from the perceptive remarks of the anonymous referees of that conference. Some of their comments are utilized verbatim in a couple of places in this paper.

As for our reconstruction of Strawson's ideas, we hope that our interpretation and (partial) reworking of his work is pretty much accurate. However, as Johnson-Laird (1990, p. 9) rightly cautions:

> [T]here is no end to the process of recovering speakers' intentions— why they chose to communicate this or that information. And a text does not talk back, and hence as its author's background assumptions fade into obscurity so its interpreters are free to project ever wider and ever more idiosyncratic readings into it.

References

Aitchison, Jean. 1997. *The Language Web: The Power and Problem of Words (The 1996 BBC Reith Lectures)*. Cambridge University Press.

Akman, Varol. 2000. Rethinking Context as a Social Construct. *Journal of Pragmatics* 32(6):743–759.

Akman, Varol and Ferda Nur Alpaslan. 1999. Strawson on Intended Meaning and Context. In P. Bouquet, L. Serafini, P. Brézillon, M. Benerecetti, and F. Castellani, eds., *Modeling and Using Context (Proceedings of the*

---

[9]While this would definitely require another, full-fledged treatment elsewhere, it must be noted that via his ingenious construct *grafting*, Derrida has argued against this possibility. With this term he refers to the process of inserting an utterance onto a context that alters its functioning (or imagining a context in which an initially meaningless sentence would have meaning). For Derrida, context is infinitely expandable. And while meaning is context-bound, context is boundless. Sympathizing with Derrida, Culler (1988, p. 148) mentions a court case, *Frigaliment vs. BNS International Sales Corp.*, where several witnesses were summoned to confirm what *chicken* meant in the chicken trade: a bird of any age or a young broiler, fryer, or roaster. The lawyers involved in the case knew that context is produced, and that since context is not saturable, a contextualization is never completed (Edmonds and Akman, 2002).

Second International and Interdisciplinary Conference, CONTEXT'99, 9–11 September 1999, Trento, Italy), Lecture Notes in Artificial Intelligence Number 1688, pages 1–14. Springer.

Akman, Varol and Mehmet Surav. 1996. Steps Toward Formalizing Context. *AI Magazine* 17(3):55–72.

Akman, Varol and Mehmet Surav. 1997. The Use of Situation Theory in Context Modeling. *Computational Intelligence* 13(3):427–438.

Austin, John Langshaw. 1976. *How to Do Things with Words (The William James Lectures Delivered at Harvard University in 1955)*. Oxford University Press. Edited by J. O. Urmson and Marina Sbisà.

Chierchia, Gennaro and Sally McConnell-Ginet. 1990. *Meaning and Grammar: An Introduction to Semantics*. MIT Press.

Culler, Jonathan. 1988. Deconstruction and the Law. In *Framing the Sign: Criticism and Its Institutions*, chap. 8. Blackwell.

Davidson, Donald. 1984. Radical Interpretation. In *Inquiries into Truth and Interpretation*, pages 125–139. Clarendon Press.

Eco, Umberto. 1984. *The Role of the Reader: Explorations in the Semiotics of Texts*. Indiana University Press.

Edmonds, Bruce and Varol Akman. 2002. Editorial: Context in Context. *Foundations of Science* 7(3):233–238.

Fetzer, Anita and Varol Akman. 2002. Contexts of Social Action: Guest Editors' Introduction. *Language and Communication* 22(4):391–402.

Grice, Paul Herbert. 1989. *Studies in the Way of Words*. Harvard University Press.

Gumperz, John. 1992. Contextualization and Understanding. In A. Duranti and C. Goodwin, eds., *Rethinking Context: Language as an Interactive Phenomenon*, pages 229–252. Cambridge University Press.

Gumperz, John. 1993. Culture and Conversational Inference. In W. Foley, ed., *The Role of Theory in Language Description*, pages 193–214. Mouton De Gruyter.

Gumperz, John. 1997. Communicative Competence. In N. Coupland and A. Jaworski, eds., *Sociolinguistics: A Reader and Coursebook*, pages 39–48. St. Martin's Press.

Harris, Wendell. 1996. *Literary Meaning: Reclaiming the Study of Literature*. Macmillan.

Heritage, John. 1984. *Garfinkel and Ethnomethodology*. Polity Press.

Jakobson, Roman. 1992. On Linguistic Aspects of Translation. In R. Schulte and J. Biguenet, eds., *Theories of Translation: An Anthology of Essays from Dryden to Derrida*, pages 144–151. University of Chicago Press.

Johnson-Laird, Philip. 1990. Introduction: What is Communication? In D. H. Mellor, ed., *Ways of Communicating: The Darwin College Lectures*, pages 1–13. Cambridge University Press.

94 / Varol Akman

Kent, Thomas. 1993. Language Philosophy, Writing, and Reading: A Conversation with Donald Davidson. *JAC Online: A Journal of Composition Theory* 13(1). http://www.cas.usf.edu/JAC/131/kent.html.

Leech, Geoffrey. 1981. *Semantics: The Study of Meaning*. Penguin. Second edition.

Perry, John. 1997. Indexicals and Demonstratives. In B. Hale and C. Wright, eds., *A Companion to the Philosophy of Language*, pages 586–612. Blackwell.

Perry, John. 1998. Indexicals, Contexts, and Unarticulated Constituents. In A. Aliseda, R. van Glabbeek, and D. Westerståhl, eds., *Computing Natural Language*, pages 1–11. CSLI Publications.

Perry, John. 2000. What are Indexicals? In *The Problem of the Essential Indexical (and Other Essays)*, pages 313–323. CSLI Publications. Expanded edition.

Récanati, Francois. 1994. Processing Models for Non-Literal Discourse. In R. Casati, B. Smith, and G. White, eds., *Philosophy and the Cognitive Sciences (Proceedings of the Sixteenth International Wittgenstein Symposium, 15–22 August 1993, Kirchberg am Wechsel, Austria)*, pages 343–356. Hoelder-Pichler-Tempsky.

Récanati, Francois. 1998. Contextual Domains. In X. Arrazola, K. Korta, and F. J. Pelletier, eds., *Discourse, Interaction, and Communication*, pages 25–36. Kluwer.

Saville-Troike, Muriel. 1989. *The Ethnography of Communication: An Introduction*. Blackwell.

Strawson, Peter Frederick. 1997a. Austin and 'Locutionary Meaning'. In *Entity and Identity (and Other Essays)*, pages 191–215. Clarendon Press.

Strawson, Peter Frederick. 1997b. Meaning and Context. In *Entity and Identity (and Other Essays)*, pages 216–231. Clarendon Press.

Ziff, Paul. 1972. What is Said. In D. Davidson and G. Harman, eds., *Semantics of Natural Language*, pages 709–721. D. Reidel.

6

Epistemological foundations for the representation of discourse context

Horacio Arló Costa

The first four sections of the paper focus on characterizing the elements that enter into the characterization of the notion of discourse context. One way of doing so is by identifying this notion with the set of commonly presupposed items. I propose a multi-agent account of context where it is essential to represent what each agent takes as being commonly presupposed, aside from what is commonly presupposed. The account requires adding dynamic features to context, in terms of the capacities of supposing that something is the case, given a current context; or in terms of the capacity of updating a context with new information. These dynamic features figure prominently in the proposed characterization.

The final sections of the paper focus on the inferential role played by the doxastic commitments induced by discourse context. I argue that these commitments do play a crucial role in understanding how agents reason defeasibly from the point of view of a given context. I discuss also some of the existing accounts, in terms of autoepistemic operators. I argue that they cannot provide a good encoding for conversational implicatures of the type Grice studied. The article offers instead an alternative account of autoepistemic inference based on an insight presented by Paul Grice in a recent addendum to his seminal article 'Logic and Conversation'.

6.1 Introduction

Utterances are evaluated in context. Suppose someone utters 'The current president lost the popular vote by almost half a million votes'. It is clear that the evaluation of such utterance depends on who is talking, as well as on the time and place of the utterance. If the speech act occurs during the year 2001 in the Unites States of America, the speaker does express a true proposition in uttering this sentence. For other places or times of utterance it is even dubious that the speaker manages to express a proposition in uttering this sentence.

The context of utterance contains also background information related to the utterance. For example, in this case it might contain: 'George Bush lost the popular vote by almost half a million votes'. Examples abound. A classic one is 'Nixon is guilty too'. An appropriate element of the context for this utterance might be 'Haldeman is guilty', and so on.

Of course, the notion of evaluation can be variously articulated. As it is clearly pointed out in **?**, there are various conceptions of the way in which evaluations of utterances depend on context. Evaluation can be understood in terms of the identification of the proposition expressed (if any) by the speaker. But evaluation can also be articulated more liberally in terms of *assertibility* conditions of the sentences used by the speaker in his utterance. And assertibility can, in turn, be a matter of probability, acceptability, justifiability or truth. We can remain neutral here regarding this issue. What concerns us is the problem of how to understand what a context of utterance is.

There is a fair amount of consensus regarding the fact that utterances are evaluated in context. There is less consensus as to what are the propositional (or sentential) components of context. Some authors propose *context sets*, which basically are non-empty sets of possible worlds satisfying certain rationality conditions **?**, **?**. This view is sometimes accompanied (but not always) by an epistemic construal of context, according to which the context set encodes the set of shared assumptions of a relevant set of agents participating in an information exchange - which, in turn, can be a simple conversation or a more sophisticated process guided by a protocol.

Other authors strongly disagree. Hans Kamp, for example, points out in **?** that 'theories of belief that identify belief with sets of possible worlds cannot differentiate them finely enough to do justice to our common understanding and use of the notion. As it is most often put: Possible worlds accounts entail that belief sentences - i.e. sentences of the form 'A believes that *s*' - are truth-invariant under substitution of

necessarily equivalents for the embedded clause s, and this does not appear to be the way in which belief sentences are actually understood.'[1] Most of the criticism comes from cognitive theories of discourse processing. Some of these theories - like *dynamics semantics***?**, or Discourse Representation Theory (DRS) **?** - focus on describing the form-meaning relation as an idealized account of the process whereby the recipient of an utterance comes to grasp the thoughts that the utterance contains. Therefore these theories privilege the study of the process by which cognitive subjects construct internal representations from syntactically specified inputs. This process is seen as an operation which always applies to a pair of structures - on the one hand the syntactic description of the sentence under consideration and on the other the internal representation, which is also syntactically specified. This internal representation functions as the *context* in which the new sentence is being interpreted. This tension - between syntactic and propositional accounts - is just one among the many stumbling blocks impeding the construction of a unified account of context.

Most of the standard theories of the common ground postulate that the right account of context in discourse should focus on an abstract representation of the presuppositions shared in common by a set of interacting agents. Moreover, the theory proposes to include in the representation not only what the agents commonly presuppose but also a representation of what they *are committed* to commonly presuppose (including all logical consequences of actual presuppositions).

I shall argue in the last section of this article that this normative account is of interest in order to capture some inferential aspects of context, namely defeasible inferences from commonly held attitudes, or defeasible inferences from the agent's take on the contents of commonly held attitudes. I shall use some insights first proposed by Paul Grice in order to differentiate these inferences from other inferential phenomena, like implicature. Nevertheless, the first part of the article will focus not on what agents are committed to presuppose, or on the agent's takes on those commitments, but on the representation of the *cognitive performances* of agents.

This presupposes an underlying distinction between normative and descriptive ideals. One possible strategy is to introduce the notion of context as a representation of the *doxastic commitments* of the speaker and hearer while a discourse unfolds. Another is to introduce contexts

---

[1]Defenders of the possible worlds account have tried to articulate a response by producing a two-dimensional construction of the so-called notion of *narrow belief*. Nevertheless this response is seldom incorporated in the standard description of context sets.

as a description of the *actual performance* of the agents involved in an information exchange.

I shall argue in favor of generalizing and extending the 'common ground' account of context - even as a normative representation of held commitments. This view proposes that contexts are constituted by the shared assumptions (presuppositions) of interacting agents. The body of shared assumptions plays, without doubt, a crucial role in understanding how utterances are evaluated in context; but an analysis of Scott Soames's characterization of presupposition **?** will suggest a different view, where shared assumptions are just one among various other relevant parameters of context. The proposed account is conceptually related to some multi-agent constructions of context common in Artificial Intelligence **?**.

This alternative view crucially requires the development of a more sophisticated conception of how contexts are updated. Most of the existing accounts (syntactic or propositional) tacitly assume a very limited view of context change. Perrault **?** has recently characterized this view in terms of what he calls the *persistence theory of belief*. We will argue that such account of context change is inadequate for many applications, and we will offer an alternative based on a proposal first presented in **?**. The consideration of various examples will suggest that an adequate account needs to be able to explain *how* to effect realistic transitions among contexts in the presence of given inputs. More importantly, we will also argue that an adequate theory of context requires an account of *what* inputs are acceptable when various stimuli are admissible. As a result we propose to incorporate the *value of information* **?** as one of the crucial components of epistemic context. The view of context that thus arises offers a parametrical account, of which both descriptive and normative accounts of context can be obtained as special cases.

6.2   The common ground as a context set

In one of his most recent articles on the representation of discourse context in **?** Stalnaker proposes the following model:

> I propose to identify a context (at a particular point in a discourse) with the body of information that is presumed, at that point, to be common to the participants in the discourse.

> [...]

> We can represent the information that defines the context in which a speech act takes place with a set of possible situations or possible worlds - the situations that are compatible with the information.

Stalnaker calls these sets of possible worlds *context sets*. Two additional constraints are imposed on them. Context sets are supposed to be non-empty sets of possible worlds, and the actual world need not be among them. We elaborate below on the meaning of these constraints.

Representing *doxastic (or epistemic) context* as a set of possible worlds is a usual strategy in formal epistemology. So, if W is the set of primitive points in the representation[2] a non-empty set $K_X \subseteq W$ can be seen as representing the strongest proposition to which some agent X (or a group of agents) bear some attitude A. If the attitude in question is, for example, belief, one can say that a proposition p is believed by agent X as long as $K_X \subseteq p$.[3]

One can define as well an operator for the attitude A. This can be done in many ways. One of these ways, which is particularly non-committal (and will be useful below) is based on the use of fixed-point equations of the following sort (considered also by Stalnaker in previous writings **?**).

(S1)] $K_X \subseteq p$ iff $K_X \subseteq \underline{A(p)}$
(S2)] $K_X \nsubseteq p$ iff $K_X \subseteq \overline{A(p)}$

These equations might be seen as establishing introspective ideals for the given attitude. For example, if the attitude A is presupposition, Stalnaker does require in **?** the introspective ideals entailed by S1-2. $A(p)$ in this case stands for 'agent X presupposes p'. According to Stalnaker: 'agents know what they are presupposing, so they presuppose that they are presupposing p if they are, and that they are not if they are not.'

Representing an attitude A via the previously suggested model is tantamount to impose strong ideals of rationality. For example, if A intends to represent the *rational beliefs of an agent X* the requirement that $K_X \neq \emptyset$ is imposed in order to express the ideal that rational agents should be logically consistent. On the other hand the fact that the actual world need not be among the worlds in the context set indicates that we are dealing with a *doxastic* rather that an *epistemic* representation - commonly held beliefs or presuppositions (unlike com-

---

[2]The set W can be understood as a set of possible worlds or situations, or a set of points in some structure, like a sigma-field or an algebra, depending on the underlying assumptions about the nature and theoretical role of W. We will not go deeper into this issue here, although the theoretical decisions as to the nature and theoretical use of W are far from being idle. For the moment we will assume that the points in W as unstructured primitives, and we will not assign to them any intended interpretation - ontological or epistemic.

[3]From now on upper case letters D, E, ..., will stand for sentences and the corresponding d, e, ..., will stand for the propositions expressed by those sentences.

monly held pieces of knowledge) might be false.

The mere fact that the A-state (belief-state, knowledge-state, presupposition-state, etc) of the agent is represented by a set of possible worlds also imposes substantial constraints. One of them is to assume that that the agent is *logically omniscient.* In other words, X believes (knows, presupposes, etc) all the logical consequences of items he actually believes (knows, presupposes, etc).

It should be rather obvious that a representation of the sort just sketched has little or no teeth as an encoding of the *epistemic performance* of agents through time. Both simple introspection and experimental work show that agents inadvertently fall into contradictions. It is also rather obvious that human agents are not logically omniscient. Representations of the sort just sketched have been used, nevertheless, in a relatively profitable manner in order to develop normative theories of rationality. For example, the set of propositions entailed by K_X can be seen as the set of propositions that X is *rationally committed* to believe (if A stands for belief). So, if X *actually* believes some finite (and consistent) set of propositions the intersection of this finite set of propositions will, in this case, be the appropriate K_X – representing the strongest proposition that X is committed to believe, given current beliefs.

Sometimes it is argued that there is a theoretical and practical link relating the idealizations of theories of rationality to the descriptive task of representing the epistemic performance of real agents (see, for example, chapter 1 of **?**). This is accomplished via a metaphor borrowed from thermodynamics. Real thermodynamic trajectories link states which might not be in thermodynamic equilibrium. Nevertheless one can gain some insight on thermodynamic processes by studying transitions between states in thermodynamic equilibrium. By the same token we can perhaps gain some insight about real epistemic transitions by studying transitions between idealized epistemic states in reflective equilibrium. The ideals of rationality encoded by theories closed under S1-2 are also usually defended in a similar manner. According to Stalnaker, for example, the idea of imposing these conditions is 'to capture the final [equilibrium] states that an agent might reach by reflecting on his beliefs and by making inferences from them and about them'. These final states must obey some intuitive conditions. An important one is that they should be *stable*, in the sense that no further conclusions can be drawn from them.

The connections between theories of rational belief and context sets require the articulation of a further step, namely clarifying the nature of the attitude that goes into context sets. Against previous accounts, this

attitude is not, according to Stalnaker, just belief (or common belief). Otherwise context sets will be determined in terms of commonly held (rational) beliefs of actors participating in a dialogue, and this might be too strong to be a viable representation of what the actors regard as 'common ground' among them. The idea is to appeal to a different attitude, namely *presuppositions*. It is not easy to determine what exactly such an attitude is. As a matter of fact, exactly determining what presuppositions are has been the object of considerable recent (and not so recent) controversy. We will verify later on that a charitable understanding of some of the proposed definitions of presupposition might require introducing enrichments in these definitions, and that those enrichments are of similar type of the kind of refinements we are about to propose for context representation. So, we will not go deeper into this issue at this juncture. It is enough for the moment to say that there is considerable agreement about the fact that the presuppositions P of an utterance U are usually considered as a species of belief. Namely any presupposition P of U is a belief that the speaker regards as *uncontroversial* **?**. What exactly 'uncontroversial' means in this characterization will be discussed later. For the moment it is enough to remark that the emerging picture after the previous analysis has at least the following ingredients: (1) discourse context should be represented in terms of the so-called 'common ground' of shared assumptions among speakers, (2) the formal counterpart of such notion (and its formal representation) is a context set, (2) such set is a non-empty set K of possible worlds. Furthermore, any proposition P entailed by K is pre-systematically understood as a species of belief that the agents deem as uncontroversial (presuppositions). The context set is therefore in logical equilibrium (all logical consequences of presupposed items are presupposed, and all logical truths are presupposed) and the context set is in introspective equilibrium (i.e. it is closed under S1-2).

How reasonable is the model presented in the previous paragraphs as a general account of (discourse) context? Ultimately we will argue that the proper account of what a context of utterance is should not be understood in terms of context sets. At least this is so if the goal is descriptive. For normative purposes we will argue that the context set account needs to be extended, adding the explicit dispositions to change view of agents engaged in dialogues. We will proceed as follows. First, we will not question assumption (1) presented above, i.e. the idea that the body of shared assumptions is the right representation of context in discourse. We will take this for granted and we will focus on (2) the adequacy and completeness of using context sets as the formal counterpart of the set of shared assumptions. We will argue that for

some applications context sets are inadequate and we will consider instead an alternative representation proposed by Richmond Thomason. This representation seems to have the advantage of explicitly adding dynamic elements into the formal representation of context. This makes the model normatively more adequate (or so we will argue). The model also makes room for weaker representations of attitudes by imposing less rationality constraints. Therefore Thomason's alternative model seems to be endowed also with more descriptive force.

Second we will discuss and then abandon (1) the idea that the body of shared assumptions is the right representation of context in discourse. We will adopt instead a slightly modified and extended version of an account first proposed by Scott Soames **?**.

As we remarked above, the use of context sets seems to borrow heavily from previous normative work in theories of rationality. One of the main goals of these theories is to construct representations of the rational commitments of agents at each instant. Much of the current work in computational linguistics (and related fields in cognitive science and artificial intelligence) seems to focus instead on the study of the cognitive representations which agents construct in response to verbal inputs. The main idea of some of these accounts (DRT, for example **?**) is to describe the form-meaning relation as an idealized account of the process whereby the recipient of an utterance comes to grasp the thoughts that utterances contain.

Even when some of the current cognitive theories appeal to idealizations, their main task seems to be primarily descriptive. Representation of context in these theories is done via finite syntactic representations. For example, in the theory presented in **?** a message carried by an incoming sentence triggers the construction of a discourse representation. This representation is sensible to syntactic considerations. Therefore two sentences carrying identical content, but distinguishable under a syntactic point of view, might produce different posterior representations (even when the prior representation coincide). For example, consider the following two sentences:

(P.) Exactly one of the ten balls is not in the bag.

(Q.) Exactly nine of the ten balls are in the bag.

P and Q carry identical content (they are true in the same set of possible worlds), but they differ syntactically. Call C the prior context and C_P and C_Q the posterior contexts after updating them with P and Q respectively. In a theory like the one presented in **?** C_P does not

coincide with C_Q.[4] Hans Kamp has recently used this example (due originally to Barbara Partee) in order to criticize the idea of representing contexts via sets of possible worlds.

The argument runs as follows. If assertion of a sentence is understood as 'a proposal to alter the context by adding the information that is the content of the assertion to the body of information that defines the context set' **?**, then C_P and C_Q should coincide. At least this is so if one accepts a principle stating that a prior context set modified by identical propositions yield a unique posterior context set. This is a reasonable principle, even when the proposition corresponding to P (Q) has an empty intersection with the prior context set C - and therefore the update in question requires a 'genuine' revision of C. But now it seems that we arrived to an unpalatable conclusion, given that the context set induced by the ulterior utterance of the following sentence:

(S.) It is under the sofa

seems to lead to posterior contexts $C_{P,S}$ and $C_{Q,S}$ which cannot receive the same interpretation.[5] Nevertheless, if C_P and C_Q coincide, $C_{P,S}$ and $C_{Q,S}$ should coincide also.

Stalnaker has responded to this argument in **?**. The gist of the rebuttal is to propose that incoming sentences update the context as a result of a two-step process. The first part of this process invokes Gricean priciples of conversation. The assertion of a sentence is, of course, a speech act, and Stalnaker argues that the mere occurrence of such act changes the context, before the agent engages in the epistemic process of changing view. Commenting on Kamp's argument against representing contexts via context sets, Stalnaker argues as follows:

> Since it is a manifestly observable fact that, in each case, a certain sentence was uttered, this fact, together with any additional information that follows from that fact, conjoined with standing background information about linguistic and speech conventions, is available to distinguish ... [C_P and C_Q] ...

What Stalnaker offers here is a two-layered theory of update. Given a prior context C and an incoming proposition expressed by a sentence S, C_S is not just the update of C by S, which can be straightforwardly

---

[4]Basically the context for C_P has a reference marker for an individual ball while C_Q lacks it.

[5]The DRT representation articulates this asymmetry as follows: since the context for C_P has a reference marker for an individual ball, while C_Q lacks it, the construction algorithm implemented by DRT recommends to take *it* as referring back to the bag when processing Q and to an individual ball when processing P.

calculated when $C \cap s \neq \emptyset$, by eliminating from C all worlds where the proposition s (correposnding to S) is false.[6] The mere utterance of S might alter C yielding a modified prior context C'. Any posterior is then calculated with respect to this modified prior.

Stalnaker's idea is to abstract away from any theory of English or from any theory of the practice of speaking English, which nevertheless can be part of the background information true in all possible worlds of the context set. Whatever the nature of the facts invoked by cognitive theories in order the explain how agents construct data structures in the presence of verbal inputs, these facts - Stalnaker argues - will be taken into account in order to distinguish the possible worlds that define the universe of possibilia from which contexts are made. Stalnaker seems to be right in pointing out that proceeding in this way does not commit him to add nothing new to the possible worlds that define context - according to his characterization in terms of context sets. When an assertion is made the set of possible worlds compatible with the relevant linguistic information might be updated by the mere fact that the speech act occurred. And this updated context set can be further updated by adding to it the content of the speech act. It should be pointed out, nevertheless, that as long as the representation of context lacks the capacity of representing pragmatic changes of the type invoked by Stalnaker, one needs an account of context change where updates with logically equivalent sentences might not produce the same output. We will elaborate on this dynamic aspect of context below.

A point needs to be made here in passing about the nature of context. Even if one assumes that the right characterization of context should proceed epistemically by specifying the body of shared assumptions of speaker and hearer, further distinctions are needed in order to narrow down the set of possible theories compatible with that view. The set of shared assumptions could just be the set of shared assumptions that speaker and hearer consciously share at same stage of a dialogue, or it could be the set of assumptions that they are committed to share (even when they are not aware of some of these commitments). The notion of commitment can be here loosely constructed in terms of the epistemic obligations of reflective and rational agents with access to the information exchanged in a dialogue. The following example can help to clarify the distinction just made. You call the cab company and ask for the number of the cab that will pick you up in the next five minutes. The dispatcher tells you:

---

[6]More about the universality of this Bayesian proposal for update below.

(D.) The number of the cab is the number 1729.

The posterior context set C'_D encodes both the pragmatic and epistemic consequences of the utterance. First the prior context might shift to a modified context C', as a consequence of the mere fact that the speech act itself takes place. Second C' is updated by d. It is reasonable to suppose that in this simple case the posterior context set C'_D entails the proposition expressed by d itself. Notice then that an agent whose presuppositions are described by the context set C'_D should be described *also* as being committed to presuppose the information carried by the following sentences:

(E.) The number of the cab is the smallest number expressible as the sum of two cubes, in two different manners.

(F.) The number of the cab is the number obtained by subtracting one to 1730.

(G.) The number of the cab is 1729 or my uncle lives in Boston.

Of course, the agent might eventually fail to realize that he is committed to presuppose that e – when he actually presupposes that d. And even when g is a logical consequence of d, the information carried by g might be deemed irrelevant.

If the theoretical goal pursued by representing context is to describe the assumptions shared by speaker and hearer, and this includes only the assumptions that they are aware of, then context sets are an inappropriate representation of context. [7] This does not mean that the

---

[7]The problem under consideration here is the so-called problem of *logical omniscience*. Once certain standard of rationality is fixed, establishing the commitments of a rational agent is unproblematic. If the agent is supposed to be logically consistent and as well as logically proficient, his view should be represented via a non-empty context set - or (syntactically) by a logically closed set of sentences strictly included in the underlying language. If the agent is supposed to be probabilistically coherent, then his degrees of belief should obey the laws of standard probability, and so on and so forth.

It is less clear how to represent the explicit beliefs of *bounded agents*. Trimming down commitments in order to capture inferences of sufficient complexity will not due. Complexity considerations are obviously agent-dependent and difficult to characterize exactly. For example, that d and e are co-extensional was perfectly obvious for the Indian mathematician Srinivasa Ramanujan - perhaps as obvious as the fact that d and f are co-extensional. When visited by a friend in the hospital, his friend told him that the number of the cab in which he traveled to visit him was 1729, a number that he considered insipid. Ramanujan responded immediately that this was not the case, mentioning the property used in e.

explicit representation of commitments is useless. I shall argue below that representing contextual commitments serves an important normative purpose in representing the inferential capacities of agents. For the moment, nevertheless, I shall pause to consider some dynamic aspects of context representation.

6.3 The conversational record and context change

It should be evident by now that when context sets are invoked in order to characterize speech acts, the dynamics of context is of foremost importance. In the previous section we considered the case of assertion. Following ideas first presented in a seminal paper published in 1978 **?** Stalnaker understands assertion as 'a proposal to alter the context by adding the information that is the content of the assertion to the body of information that defines the context set'.[8] In other words, *the goal of asserting* a proposition is to update the current context. But as Stalnaker clearly explains assertion as a speech act cannot be *identified* with the update operator:

> I should emphasize that I do not propose this a DEFINITION of assertion, but only as a claim about one effect which assertions have, and are intended to have - an effect that should be a component, or a consequence, of an adequate definition. There are several reasons why one cannot define assertion in terms of this effect alone. One reason is that other speech acts, like making suppositions, have and are intended to have the same effect. A second reason is that there may be various indirect, even nonlinguistic, means of accomplishing the same effect which I would not want to call assertions. A third reason is that the proposed essential effect makes reference to another speech act - the rejection of an assertion, which presumably cannot be explained independently of assertion.

In fact, there are a variety of non-verbal acts capable of effecting transitions of currently held attitudes. Thomason mentions the well-known example of Herod when he caused John the Baptist's head to be brought in a platter. Of course, Herod did not assert anything, but his action caused an update in the body of presumptions. On the other hand, epistemology commonly studies *mental acts*, like acceptance, whose (constitutive) point is also effecting transitions of epis-

---

One of the possible manners of dealing with this is to derive explicit beliefs in a principled manner from other theoretical considerations, which could be probabilistic **?** or algorithmic **?**. For example, the first strategy would suggest representing explicit beliefs in terms of high probability. No principled criteria of this sort will be invoked here, but we will make room for representations of this sort by weakening the context set approach.

[8]The quote is from **?**.

temic states. Moreover, even if the point of asserting P is to produce a change of view in the current context, the change in question might not be straightforwardly construed as a function of the content of P and the body of information that constitutes the context set. We just saw the current context C might change as a result of merely asserting P.

But even if speech acts, like assertion, cannot be fully characterized in terms of epistemic transitions, it is undeniable that one can gain interesting insights on those acts by studying the epistemic transitions they purport to effect. Even if this is so, most of the current theories of context refrain from adding specific dynamic parameters to context sets. Two underlying assumptions can perhaps be invoked as an explanation of this fact. The first assumption is based on the idea that all needed changes can be construed as incremental additions to the current context set. In other words, all represented transitions are construed as changes where the set of possible worlds constituting the current context C has non-empty intersection with the input proposition p. Let's call such changes *additions*. The second assumption suggests how to perform additions. The idea is to add p to a context C by just eliminating from C all possible worlds where P is false. We intend to argue here that both assumptions are based either on misconceptions or on simplifications, which are unduly restrictive. In a recent article Perrault **?** clearly articulated the first assumption, which he calls the *persistence theory of belief.*

> Speech acts reveal certain aspects of the speaker's mental state and cause changes in the state of the hearer(s) that are based on their perception of the state of the speaker. An agent's beliefs after an utterance, for example, will in general depend on his beliefs before it, as well as on its content. Ideally, one would like to have a theory in which it is possible for ones agent's beliefs, say, to change according to how strongly he believed something before the utterance, as well as on how much he believes what the speaker says. I cannot give such an account in detail, so I will rely on something simpler. I assume what might be called a *persistence theory of belief*: that all beliefs persists and that new ones are adopted as a result of observing external facts, provided that they do not conflict with new ones. In particular, I assume that an agent will adopt the beliefs he believes another agent has, as long as those do not contradict his existing beliefs.

Perrault is aware of the fact that adopting persistence is a gross oversimplification. There are various cases that can only be analyzed in terms of belief-contravening changes of view. The contextual evaluation of conditional sentences is an obvious case, but examples abound. John

and Terry might be looking for Paul. They might take for granted that Paul is somewhere on campus (in the city of Pittsburgh). Nevertheless, if they meet Al and he tells them that Paul is in NYC (or if Al tells them something that presupposes so), they might face a non-persistent change of presumptions - as long as they consider Al as a reliable oracle. In order to input consistently the information provided by Al some information has to be *contracted* from the prior context.[9]

The foundational and mathematical aspects of a theory of contraction (and revision) of belief has been worked out by epistemologists, computer scientists and decision theorists during the last 20 to 30 years - see, for example, **?**, **?**, **?** for general background on recent work. Parallel work has been done in analyzing the theory of *supposition* that goes into the evaluation of conditionals. Much of this theory can be immediately applied in order to analyze context change in the domain we are studying. Nevertheless the application calls for a careful use of some of the tools already developed, and in some cases it seems that suitable modifications and extensions are needed. Some preliminary observations concerning these issues will be made below. Nevertheless, our main point in this piece is to rethink what goes into the theoretical notion of context. And the relevance of the issue at hand for our main theme is that, once the need for representing non trivial additions is recognized, the notion of context set is too poor as a general foundational device for representing context. Additional dynamic parameters need to be added.

This point is not completely new in the standard literature. Richmond Thomason elaborated it in **?**. The theoretical device used by Thomason in order to represent context is different from Stalnaker's context sets in many ways, as the following paragraphs make evident:

> [...] the most important component of the conversational record is a structure **P** that determines the *presumptions*: the things that are supposed, or established, at a given stage of the conversation. I'll say that

---

[9]Stalnaker distinguishes between DEFECTIVE and NONDEFECTIVE contexts in **?**. A nondefective context is one in which the presuppositions of the various participants in the conversation are all the same. Stalnaker thinks that defective contexts have a kind of intrinsic instability, and that therefore will tend to adjust to the equilibrium position of a nondefective context. Of course, non-persistent context changes presuppose defective contexts. Stalnaker assumes that in the normal case contexts will be close enough to being nondefective. The assumption seems too strong. Perhaps Stalnaker uses the assumption because he thinks that defective contexts can lead to communication failure. This could happen in extreme cases but it seems also true that there is a large set of cases where the process of adjusting towards the equilibrium position of a nondefective context can be accomplished via non-persistent updates, without threatening communication.

$\mathbf{P} \vdash p$ when $\mathbf{P}$ yields the conclusion that p. These presumptions are perhaps best thought of as a kind of shared memory or common database that the participants construct for the purposes of the conversation.

Of course, presumptions will in general be modified in the course of a conversation. I'll suppose that for each proposition p there is an update operator Ap on the presumptions that gives the result of updating the presumption structure by adding the proposition p. Ap$\mathbf{P} \vdash p$ and in many cases Ap$\mathbf{P} \vdash q$ if $\mathbf{P} \vdash q$. However, update is a nonmonotonic operation, so that we can have $\mathbf{P} \vdash q$ but Ap$\mathbf{P} \nvdash q$; also we can have Ap$\mathbf{P} \vdash r$ but Ap $\wedge$ q$\mathbf{P} \nvdash r$, where '$\wedge$' is propositional conjunction. Since the conversational record is public, the update operator must be public also. I assume that the reasoning mechanisms that operate here are essentially the same as those that operate in the suppositional or conditional reasoning.

This previous account of context (aside from explicitly including dynamical elements) is more liberal than the one given via context sets.[10] The nature of the structure $\mathbf{P}$ is left unspecified. Notice that, in particular, the behavior of the structure need not be determined by what is entailed by a set of possible worlds. For example, the behavior of a permissible instance of $\mathbf{P}$ might depend on what propositions are members of a set of sets of possible worlds S. In other words, SP might just be the set of presumed propositions, in such a way that $\mathbf{P} \vdash p$ if and only if $p \in SP$. When this is the case we can say that $\mathbf{P}$ is *determined by a presupposition set* SP. This representational structure is capable of circumventing some of the problems mentioned at the end of the previous section. For example, if $\mathbf{P}$ is determined by a presupposition set SP and $\mathbf{P} \vdash p$, there is no need that $\mathbf{P} \vdash r$, when $p \subseteq r$. This filters the 'irrelevant' case mentioned at the end of the previous section. In this case we had $\mathbf{P} \vdash d$ and the fact that $d \subseteq g$ entailed the unintuitive conclusion $\mathbf{P} \vdash g$. If we use context sets in order to determine the behavior of $\mathbf{P}$-structures this consequence is inescapable. In order to see this it is useful to notice that context sets are a limit case of structures of type $\mathbf{P}$. Let $\mathbf{C}$ be a set of possible worlds. Then if $\mathbf{P} \vdash q$ if and only if $\mathbf{C} \subseteq q$ we have a limit case where all presumed propositions are entailed by their infinite intersection $\mathbf{C}$. In this limit case we can say that $\mathbf{P}$ is *determined by a context set* $\mathbf{C}$. Now, of course, when a $\mathbf{P}$-structure is determined by a context set, it is easy to see that if $\mathbf{P} \vdash d$ and $d \subseteq g$,

---

[10]It is important to keep in mind that the P-structures just introduced focus on describing occurring or explicit attitudes, not on inference. Perhaps an agent can presuppose that X is a lawyer and that Y is also a lawyer. He might also presuppose that both X and Y are lawyers, but the P-structures do not require universal closure under conjunctions. After section five I shall consider the *epistemic commitments* contracted by finite doxastic representations of the kind we are studying here.

$\mathbf{P} \vdash g.$[11]

So, *some* of the problematic cases considered in the last section can be circumvented by using **P**-structures. It is easy to see though that not all examples are so easily handled by adopting **P**-structures. In fact, if d, e and f are the propositions expressed by D, E and F, then it is clear that d, e and f are *the same* proposition. So, nothing forces us to legislate that $Ad\mathbf{P} \vdash g$, but $Ad\mathbf{P} \vdash e$ is unavoidably determined by the fact that **P**-structures are updated with *propositions*.

Most of the standard theories of belief change allow for finer-grained representations. In fact, most of these theories study how databases (eventually composed by sets of sentences) change when *sentences* are added, preserving consistency. We can call these theories *sentential theories of update*. Nevertheless, most axiomatizations of these sentential functions undo this freedom by adopting a postulate usually called[12] the Principle of Irrelevance of Syntax. The principle legislates that updates prompted by co-extensional sentences are identical. This principle is adopted in virtue of various reasons. One reason is mathematical simplicity. Another, more principled, reason is based on the fact that most of the existing theories of belief change are *normative theories*, describing how the doxastic commitments of rational agents are updated over time.

Some of the cognitive theories of discourse processing, like DRT, focus on syntactic representations of context. Hans Kamp, for example, points out in **?** that 'it is tempting (and, I think, up to a point legitimate) to see DRT as a providing a model of the process by which the recipient of a discourse acquires new beliefs as he takes in successive sentences.' Now, we already know that Kamp is a critic of the context set model, and this is evident here too. After proposing examples of the type we considered above, Kamp refuses to analyze belief as a relation between a believer and a set of possible worlds. Kamp adopts instead a syntactic model according to which (roughly) 'beliefs are identified with sentences of a natural language such as English, while recognizing the need to explicate the intentionality of that natural language, presumably by providing a truth- or model-theoretic semantics for it.'

A detailed comparison of P-structures and DRT is beyond the scope of this piece. Such a task requires making explicit the general principles of doxastic change tacitly used in rich accounts of discourse processing (like the one offered in **?**. It should be mentioned in passing nevertheless

---

[11]The behavior of **P**-structures is reminiscent of the so-called *neighborhood semantics* for modalities defended by Scott and Montague. See Arló-Costa (forthcoming) for a direct application of this type of semantics to epistemic logic.

[12]Specially among computer scientists. See, for example, **?**.

that the Principle of Irrelevance of Syntax is not among the general principles governing doxastic change in this theory. Given the centrality of this principle in almost all contemporary theories of belief change, it is fair to assume that the format of the theory of change arising from the use of the construction algorithm in **?** is today mostly unknown.

6.3.1 Success and the act of supposing

Aside from the previous considerations about update, it is important to see that the only substantial property imposed on A is the property that we can call Success: $Ap\mathbf{P} \vdash p$. Success requires that when an agent represented by a structure $\mathbf{P}$ updates his view with a proposition p, the resulting updated structure yields the conclusion that p.[13]. Some minimal considerations about the notions of supposition and belief revision (with external inputs) can help to put in perspective this and other properties of A.[14]

Notice that changes in view might occur as the result of processing inputs which might have different epistemic origins. Perrault explicitly assumed that 'new beliefs (presumptions) are adopted as a result of observing external facts'. But inputs can, of course, have a purely internal origin. This is the case of changes of view done 'for the sake of the argument', or as a result of *supposing* an item. In other words, supposition can be understood as a species of update. This is indeed the way in which supposition is usually construed in many branches of Bayesian epistemology.

Supposition is the mental act needed in order to evaluate conditionals. Consider John who might face the task of considering the acceptability of the conditional 'If Paul is in NYC, the he will return on Monday'. Even if John firmly believes at the moment that Paul is somewhere in the city of Pittsburgh, he can evaluate the conditional by changing hypothetically his mind. He can do so by considering for the sake of the argument that Paul is in NYC. The construction of this hypothetical state will be a function of the background information John currently holds and the hypothetical input he is entertaining.

Success is an appropriate (if not a constitutive) principle that any reasonable suppositional operation should satisfy. It is less clear that updates caused by external inputs should be constrained by Success. If Al tells John that Paul is in NYC this event will trigger a process of belief revision. John will change his mind after receiving the report,

---

[13]Sentential theories of update would implement: $AP\mathbf{P} \vdash p$

[14]For the sake of terminological clarity we use the term 'update' in order to refer to any change of view, hypothetical or not. 'Belief revision' will be used in order to refer to updates with external inputs.

but after changing his mind John need not believe that Paul is in NYC. His current belief to the effect that Paul is in Pittsburgh might be so well *entrenched* in his prior view, and his trust in Al as a reliable oracle so flimsy, that John might decide to stay put. Therefore his view about Paul's location might not change after receiving Al's report. Of course, John's posterior view will be different. Among other things John will firmly believe that he received a report from an unreliable source about Paul's whereabouts. But notice that the mental act of supposing does not seem to involve this epistemic freedom. Supposing that p necessitates the implementation of a hypothetical update entailing p. Otherwise one might say that an attempt to suppose that Paul is in NYC failed (the scenario might be unconceivable, given prior beliefs).

Thomason seems to have the notion of supposition in mind when he proposes to constraint A by Success. And we just argued that, for this theoretical application, the postulate seems well motivated. Nevertheless, we will argue in the next section that some applications might call for the addition of other update operators for each agent, which need not obey success. Most of the recent literature on belief revision (with external inputs) has adopted the Success postulate, but there is an increasing consensus in the field towards developing more sophisticated models, which are usually called *non-prioritized*.[15] We will comment on this issue in the coming section.

6.3.2   Presuppositions and update

In the previous section we argued in favor of a model of context of the type sketched by Thomason in **?**. The model is at the same time weaker and more expressive than models in terms of context sets. Weaker because it allows for finite of explicit attitudes. More expressive because it explicitly contains dynamic parameters.

Of course, as we saw in the previous section, more work seems to be needed in order to complete the picture first offered by Thomason. For example, it would be nice to have a better understanding of the nature of the dynamic parameters included in the model. More work seems to be needed as well in order to understand the nature of the attitude that goes into the conversational record. Stalnaker talks about presuppositions, Thomason uses the term 'presumptions', Perrault and others directly refer to beliefs. In this section we will try to tackle some of these issues. Our starting point will be a characterization of presupposition offered by the philosopher Scott Soames in **?**.

(P) An utterance U presupposes p (at t) if and only of one can reason-

---

[15]See **?** for an introduction to work in this area.

ably infer that the speaker accepts p and regards it as uncontroversial, either because:

a. S thinks that it is already part of the conversational record, or because

b. S thinks that the audience is prepared to add it, without objection, to the context against which U is evaluated.

So, while John is looking for Paul, he might find Mary, and she might volunteer the following piece of information.

(S) Paul is not on campus, he went to the airport in order to pick up his sister who is flying from Boston.

According to (P) the utterance of (S) presupposes the fact that Paul has a sister, as long as Mary believes so and the audience 'is prepared to add this fact, without objection, against the context against U is evaluated'. Various issues need immediate clarification.

First let's just assume, for the purposes of considering the example, that the audience is constituted by John. Secondly let's focus on the dynamic aspects of the definition. Notice that the relatively terse presentation of clause (b) hides a considerable degree of complexity.

Let's first tackle this intuitively by considering a case where the act of uttering (S) does presuppose that Paul has a sister (let's call this proposition p). Notice that Mary might believe that John does not presuppose p. She might believe also that John believes that Paul does not have a sister. Moreover she might think that John thinks that it is part of the common ground that Paul does not have a sister. This will not stop Mary from presupposing p as long as she thinks that John will be willing to take her words at face value and revise his views as a result.[16]

If John considers Mary a reliable source of information, the utterance of (S) will generate an immediate change in John's view about the presupposition set. If John thinks, before the utterance, that it is part of the common ground that Paul does not have a sister, the mere utterance of (S) will cause a change in John's take on the presupposition set. After the utterance John cannot think any longer that it is commonly presumed that Paul does not have a sister. This will lead to

---

[16]This assumes that Soames' idea of characterizing presupposed items in terms of what is uncontroversial for the audience should not be equated with the idea of characterizing presuppositions as uninformative items - for the audience. More about this below.

a contraction of his take on the presupposition set. I.e. John will have to move to a position of doubt considering p. Of course, this contraction might be followed by the addition of p, after the intermediate state of epistemic suspense. The compound process of contracting and then (eventually) adding leads to a revision with p.

Now, notice that what counts here in order to determine whether p is presupposed by uttering s is Mary's expectation as to whether John will actually revise his view. John may very well refuse to open his mind in order to accommodate p. This might be due to different reasons. For example, John can consider Mary an unreliable source of information.

Whether the utterance carries or not the presupposition largely depends of how the expectations of agents interact. For example, Mary might take for granted that Paul has a sister and she might not know whether John has any views about this issue. In this scenario she can utter (S) and she can presuppose p while doing so. In other words, she might think that either John knows about p or that he will unproblematically accommodate p in his view in order to continue the conversation. But, of course, her expectations might miscarry and John might refuse to do so. In this case the conversation can continue as follows. John might say:

Paul's sister? I did not know he had a sister. Are you sure we are talking about the same person?

Of course, the fact that John does not respond as Mary expected, does not conflict with the fact that her utterance did carry the presupposition that p.

We will consider below similar examples as well as some counterexamples. But before doing this let's clarify the formal background presupposed in the analysis. For the sake of terminological simplicity we will keep unchanged the notation used above, when possible.

> An *interactive structure* M is a triple $\langle I, S, U \rangle$, where I is a set of n agents (which at each instant can be partitioned into a speaker s and $n - 1$ members of the audience); S is a collection of structures $\{\mathbf{P}, P^1, ..., P^n\}$, and U is a collection of update operators $\{A, A^1, ..., A^n\}$. As before the propositions entailed by the structure $\mathbf{P}$ constitute the body of commonly presumed information. Each structure P^i, with i ranging over I, encodes the propositions that agent i thinks are commonly presumed. Finally A is an update operator on $\mathbf{P}$, which, as before, is successful and non-monotonic. Each A^i, on the other hand, is an update operator on the structure P^i, mapping the prior P^i and non-empty proposition into posterior structures $A^i \mathrm{p} P^i$, for arbitrary (non-empty) propositions p. None of the update operators for the n agents

are supposed to obey Success, and all of them are non-monotonic, in the sense defined above. In addition we assume the following properties (a structure is called consistent if and only if it is not the case that it entails all propositions).[17]

(Preservation) If for all p such that $P^i \vdash p$, $p \cap q \neq \emptyset$, then $A^i \mathrm{q} P^i = \{r = p \cap q: P^i \vdash p\}$

(Consistency preservation) $A^i \mathrm{p} P^i$ is consistent.

We are primarily interested in *dialogues* where I has cardinality two. The letters 's' and 'h' will be used in order to represent speaker and hearer in a dialogue. Of course, the information provided by updates of P^h, is assumed to be provided by the speaker s. Now we can present Soames' view as follows:

(IS) An utterance U presupposes p (at t) if and only if one can reasonably infer that the speaker accepts p and regards it as uncontroversial, either because:

a. S thinks that $\mathbf{P} \vdash p$, or because

b. S thinks that the audience is disposed to revise its view (successfully) with p (i.e. the speaker believes that $A^h \mathrm{p} P^h \vdash p$).

We should pause here for a moment in order to sum up some of the main features of the analysis so far. First, the reader probably noticed that although the structure $\mathbf{P}$ figures in the definition of interactive structures, $\mathbf{P}$ does not play any role in (IS). Clause (a) uses the $\mathbf{P}$-structure, but (a) can be easily reformulated without it. In fact, the role of P^s is exactly to encode the propositions that the speaker thinks are entailed by $\mathbf{P}$. So, clause (a) can be rewritten as follows:

a. $P^s \vdash p$.

To be sure, a group of agents sometimes directly engages in the process of updating their presuppositions. This can happen in different manners. A simple one occurs when the group consciously engages in the task of building presuppositions step by step. Teaching, joint research or any other complex group activity might require being sure that agents are in epistemic control of the nature of their agreements. But in less regimented situations agents might only have beliefs about

---

[17]Similar principles hold for sentential updates. Such updates will be used in some of the following sections.

116 / Horacio Arló Costa

commonly held attitudes. When two agents initiate a dialogue they might in fact share a large body of beliefs, but they might not know that this is the case. As a matter of fact the point of a dialogue is, at least in part, to verify which is the nature of the agreements and disagreements among the participants in the dialogue. So, it seems that a general description of a dialogue requires postulating not only structures in order to describe the agreements among agents, but it also requires postulating structures describing the agent's beliefs about the nature of the agreements. This is the role played by the P^i structures in the model. As we explained above, the updates of these structures is supposed to model not a process of supposition, but a process of belief change with external inputs, which need not obey Success.

A detailed account of the type of revision of P^i-structures requires more elaboration. For example, an analysis of the reliability of sources of information needs to be entered into the model. And what is needed is not only an account of the reliability of the speaker, but also an account of the expectations that agents have about the reliability of the speaker.[18]

We will not go deeper into this issue here, given the foundational nature of this paper. It should be said, nevertheless, that the update process required by clause (b) of (IS) is presented by Soames as a *cognitive decision*, where the agent makes up his mind about accepting, or not, a piece of information. Most of the existing theories of belief change cannot represent this process. This is in part due to the fact that these theories have focused on studying the process of changing view, *after* it is settled that an input will be added to the current view (preserving consistency). In other words, most of the existing theories of belief change accept the postulate of Success and focus on the nature of the process *after* a decision has been made to input information.[19] More sophisticated theories, like the one presented by Isaac Levi in **?** and **?** give a more comprehensive picture. These theories understand belief change

---

[18]Some recent work has been done in this direction in a probabilistic tradition. An article in this vein, available to members of the context community is Bovens and Hartmann (2001).

[19]The imposition of the postulate of Success in many contemporary theories of belief change was probably motivated by a methodological decision made in order to divide theoretical labor. The idea in **?**, for example, is that the study of how to accommodate information inconsistent with a view, while preserving consistency, is complex enough in order to study it separately. These theories therefore focused in studying the process that follows the decision of imputing belief-contravening pieces of information, leaving aside the complementary process of deciding *what* to accept. Theories of acceptance focus on this second aspect of belief change. It seems that the correct articulation of (IS) requires attending to *both* processes.

as a decision, where part of the problem is to decide *what* to accept, aside from the problem of *how* to accept information in conflict with the current view, in case that this were needed. In order to implement theories of this sort one needs to add a new important parameter in order to characterize context: *value*. In this case we need some measure of the *value of information*, aside from the standard doxastic parameters. Levi, for example, proposes in **?** an account of contraction as an optimal cognitive decision where the goal is to minimize losses of informational value. The theory also has the resources needed to characterize the process of acceptance (i.e. the resources needed to decide *what* to come to believe). Once a theory of this sort is implemented we have a notion of context including a full-fledged epistemological theory as one of its components. The route that led us here is simple. Dynamic parameters are needed in order to specify contexts. And those parameters require the use of a theory of acceptance, i.e. a theory capable of implementing cognitive decisions. We will not go further here concerning this issue, but ultimately the correct specification of context seems to require, at least, this level of complexity.

The evaluation of clause (b) of (IS) also involves suppositional aspects that is worth analyzing in some detail. What is required is a process by which the nature of an update (made by the hearer) is estimated from the point of view of the speaker. We have not considered so far how exactly this process takes place. Perhaps the speaker *simulates* the hearer's body of information (as well as an update of it) in order to form his belief about the prospects of success of the operation in question. This might involve an elaborate process of supposing. The speaker supposes that he is a different agent and he also supposes that an epistemic input is provided to this hypothetical body of belief. Then he can estimate what he would do in case of being in this epistemic situation. Ultimately what transpires in clause (b) is the evaluation of complex conditionals of the form 'If I were the hearer, I would do this and that'. More precisely: 'If I were the hearer, and I were facing a process of belief revision with external input p, then I would decide to accept the input and perform the revision (or I would decide not to accept the input, etc)'. The previous paragraph focused on the complexities associated with this decision. Now, we will pay closer attention to the role of supposition on the part of the speaker.

One simple way of tackling this problem is to postulate a further structure $P^{s,h}$ representing the view that the speaker has about hearer's point of view and an update operator $A^{s,h}$ that the speaker thinks the hearer has. Update operators on this structure are supposed to obey

identical properties than the A^i operators - in particular they are not constrained by Success. Then we can reformulate clause (b) as follows:

(b) S thinks that the audience is disposed to revise its view (successfully) with p (i.e. $A^{s,h}\mathrm{P}P^{s,h} \vdash p$).

Here we added an element previously considered, namely the update operator is sentential, mapping pairs of sentences and propositions to propositions. And, as we discussed above, the operator in question need not obey the principle of Irrelevance of Syntax.

The nature of the kind of suppositional reasoning involved in the speaker's performance is still slightly masked by this representation. More can be done in order to make the suppositional aspect of the reasoning required by (b) transparent. In order to do so we need to represent update operators in the object language. Computational theories of update do this routinely. The idea is to work with structures that are finitely axiomatizable, in the sense that given P^i, there is a sentence X, such that the proposition x determines P^i. Then we have that $x \subseteq p$ if and only if $P^i \vdash p$.[20]

Now the formula $X_h * P$ can be used in order to represent the hearer's update operator in the object language. $X_h * P \rightarrow P$ indicates that the operation is successful, where $\rightarrow$ is material implication. $|X_h * P|$ represents the proposition corresponding to $X_h * P$. Let now S^s be a suppositional operator for the speaker *obeying Success*. And, of course, '*' is not constrained by Success. We have now all the elements on order to represent the suppositional elements of clause (b) of (IS):

(Suppositional representation of b) S thinks that the audience is disposed to revise its view (successfully) with P (i.e. $S^s(X_h * P) \vdash |X_h * P \rightarrow P|$).

The idea of the representation is that if the speaker supposes that the hearer considers an update with P, then he expects this update to be successful. Thomason's idea that the main ingredients of the reasoning required to characterize context are of suppositional nature is basically correct. Nevertheless, the exact nature of the required reasoning is a bit elusive. The speaker needs to engage in an involved type of suppositional reasoning, entertaining hypotheses about whether the agents in the audience will or not decide to accept certain pieces of

---

[20]Weaker assumptions can be made here when the structures are finite. In those cases we can just take a conjunction of sentences representing each entailed proposition as the syntactic representative of the structure.

information.

6.4   What is a context of utterance?

For Thomason the body of shared assumptions is the most important
component of the conversational record. This is not so in the theory
sketched above. Ultimately it is the body of speaker's presuppositions
the part of the record that plays a crucial role in understanding the way
in which a conversation unfolds. And, following our reconstruction of
Soames, in order to determine what goes into the body of the speaker's
presuppositions we only need to appeal to: (1) the speaker's beliefs
about what is entailed by the shared assumptions, (2) the suppositions
of the speaker about how other agents would change their views as a
response to verbal stimuli.

   Soames' theory has a dual aspect. On the one hand, it can be seen
as a characterization of the speaker's presuppositions. On the other
hand it can be seen as distilling the presuppositions of utterances or
the presuppositions carried by sentences. In the former case we do not
need an observer making inferences about the epistemic states of the
speaker. Remember that the gist of Soames' idea is that 'an utterance
U presupposes p if and only if one can reasonably infer from U that the
speaker accepts p and regards is as uncontroversial'. But, we can easily
distill a notion of speaker's presupposition from (P), by simply getting
rid of this external observer:

> (Ps) The speaker S presupposes p (at t) by uttering U if and only if S
> accepts p (at t) and he regards p as uncontroversial, either because:
>
> a. S thinks that it is already part of the conversational record, or
> because
>
> b. S thinks that the hearer is prepared to add it, without objection, to
> the context against which U is evaluated.

First we can define the set of propositions that the speaker thinks
belong to the common ground.

$$PSs = \{p\colon P^s \vdash p\}.$$

   In addition we have an important set of sentences ACs collecting
the propositions that the speaker thinks the hearer will be willing to
accommodate in his view. This can be defined by using our doxastic
representation of clause (b) in Soames' definition.

$$ACs = \{p\colon A^{s,h}\mathrm{P}P^{s,h} \vdash p).\}.$$

The presuppositions of the speaker are therefore defined as $Prs = \{$s believes p and $p \in PSs \cup ACs\}$. This is, of course, just a set of propositions, but there are a fair amount of dynamic and static parameters that go into its definition.

Is there a well-defined notion of what is presupposed by an *utterance type* or even by a *sentence type*, over and above what particular agents can presuppose while participating in a dialogue? Soames' definitional strategy clearly shows how one can proceed towards introducing such notions. The first step is to postulate a 'neutral' or 'prototypical' observer making 'reasonable' inferences. The second step is to postulate 'normal conditions of utterance' and define: 'A sentence S presupposes p if and only if normal utterances of S presuppose p'. Now what is presupposed by a sentence depends on what is presupposed *by an utterance*, not by what is presupposed by the agent uttering the corresponding sentence. And what is presupposed by an utterance depends on what a prototypical 'reasonable' observer concludes from evidence provided by verbal behavior. Various levels of normality are compressed in this definition. So, one should not be surprised by its methodological fragility.

There is a important tradition in pragmatics recommending rigorous terminology when it comes to consider the roles of agents in information exchanges. For example, one should resist the temptation of simplifying and saying that sentences express propositions. As Searle points out: 'I do not know how sentences could perform acts of this kind' **?**. By the same token, strictly speaking, sentences do not carry expectations, or, for that matter, they do not presuppose anything. The same applies to utterances. Agents presuppose or suppose or believe. Sentences or acts of uttering don't. One can, for sure, say that an act presupposes the satisfaction of some condition, but this is not the epistemological use of the term presupposition. Presupposing is here understood (and we argue it should be understood) as an epistemic attitude. And only entities for which agenthood can be appropriately predicated are capable of having attitudes towards propositions.

To be sure, if certain types of presuppositions are sufficiently robust, in the sense that agents tend to presuppose them every time that certain form of words is uttered in relevant conditions, then one can legitimately attach the content of the attitude to the utterance as a default. The methodological robustness of this move will be, of course, dependent of the robustness of the association between utterances and epistemic states. Engaging in the specification of these default rules might be an exceedingly complex task. But such task presupposes a previous epistemological analysis of the various mental acts of presupposing occurring while a dialogue unfolds. This article is mainly

concerned with the latter issue.

What is presupposed by utterances can be articulated in terms of what is commonly presupposed by agents. This is the gist of the idea of the so-called common-ground theories. But one of the points of Soames' analysis is to argue that this strategy fails for many types of presuppositions (like informative presuppositions). This typically happens in out-of-equilibrium scenarios where coordination misfires, or when the agents are not in epistemic control of the body of shared assumptions. But most common dialogues are of this sort. In fact, most dialogues are in part used as instruments to discover what is commonly assumed. And in cases of this sort it seems that an analysis of context of utterance should be multi-agent in nature. In order to predict verbal behavior, or even explain it, it seems that one should appeal to the presupposition sets of the various agents engaged in the dialogue and study how these sets are formed and how they change over time as well as the conditions of compatibility among them. Under this point of view we wholeheartedly agree with the methodological stance adopted in recent work in computer science where the idea is to formalize context by studying the various viewpoints of agents, plus conditions of compatibility among them **?**.

Now, in spite of the previous defense and further articulation of Soames' theory presented above, it should be said that the theory has been recently criticized and that, therefore, some of these criticisms can easily be used in order to criticize the view of context outlined in the previous two sections. We will consider some of these arguments in the following paragraphs. Let's start with and example originally due to Karttunen **?**:

(K) We regret that children cannot accompany their parents to the commencement exercises.

The speaker can perfectly well presuppose in uttering K that children cannot accompany their parents to the commencement exercises (a proposition which we will abbreviate by k). This might be so even if the speaker is not sure whether k is a shared assumption. The speaker might utter (K) every year, and so far nobody objected. So, although every year the audience changes, the speaker can safely suppose that this year's audience will be willing to accommodate k in their views. According to Soames then k is among the speaker's presuppositions.

In a recent paper **?** Gauker objected as follows: 'Soames' own examples show that he does not think 'uncontroversial' means *uninformative*,

but is there any other sense in which the presupposition of (K) has to be uncontroversial? One can well imagine a parental revolt, in which the parents insist that the children must be admitted and bring them into the auditorium whether the authorities permit it or not. In that case it will be hard to find any sense in which the presupposition of K is uncontroversial.'

There are two possible scenarios we can entertain here. Either the speaker is aware of the existence of the parental revolt at the moment of utterance or not. If he is not, what counts is what he thinks and supposes about the situation. And, since by assumption he knows nothing about the parental discontent, he might assume by default that nothing anomalous happens. Therefore he is entitled to suppose that the audience will be willing to modify their views with k. As we explained above, ultimately what the speaker evaluates is the outcome of a decision the audience faces. And this decision might involve weighting values of different sort. Here the values are not only cognitive in nature. The audience might weight different types of costs and loses: how costly might be to challenge authority, how much is lost by not bringing their children, how great is the reward of participating in the ceremony peacefully, etc. All this can perfectly be included in the deliberation on whether to accept k or not. As long as the speaker does not have evidence to the contrary, and in the presence of previous experience, he is entitled to conclude that the audience will be willing to accept k. And if this is so, it seems perfectly correct to say that he presupposed k in uttering (K). Even if as a matter of fact the parental revolt is in progress, there is a perfectly defined sense in which the speaker might think that the audience will find k uncontroversial. It only happens that in doing so he is misjudging the values and doxastic dispositions of this year's audience.

In the other relevant scenario the speaker has reasons to believe that a parental revolt is in the making. But then in uttering (K) it seems reasonable to say that he is not taking k for granted. He might utter (K) without taking k for granted as the last attempt to prevent a conflict. If uttering (K) felicitously requires the speaker presupposing k, then this will be a situation where this does not happen. (K) in this scenario can be seen as an indirect speech act destined to transmit (as politely as possible) a command. And whether the speaker has serious doubts about whether the command will be obeyed or not. Of course Soames' theory can handle this scenario as well, due to the fact that in this case the speaker has enough evidence to falsify clause (b) in (IS).

What is crucial in Soames' definition is not whether the information carried by k is or not uncontroversial (in some objective sense),

or the *facts* as to whether there is a revolt in the making. What is crucial is whether the speaker *thinks* (is justified in thinking) that k is uncontroversial.

From the consideration of cases like the one depicted by the first scenario presented above, Soames concludes that (in those cases) 'a speaker's utterance presupposes a proposition, even thought the speaker himself does not presuppose it in the sense' defined by assuming-to-be-part-of-shared-assumptions notion.' In other words, the structure *Pr*s rather than the structure **P** seems to be the right representation of the context of utterance of (K).

We do agree with critics of the common-ground theory that Karttunen's example seems to show that 'the distinguishing feature of the members of the context cannot be that they are already shared assumptions at the time of utterance' [**?**, page 162]. Nevertheless, this criticism does not seem to grant a fundamental shift in the methodological analysis of context. Some authors, like Gauker in **?**, for example, have concluded that such a shift is indeed needed. The alternative is the postulation of an ontological or objective view of context, according to which the context of utterance is made up of a set of propositions that are objectively given as those propositions that are relevant to the current conversation.

Most of the arguments offered by Gauker in favor of the objectivity of context seem to be based upon observations about asymmetries between the body of speaker's presuppositions and the body of hearer presuppositions. For example, Gauker makes the following observations concerning Karttunen's example:

> ' in light of the phenomenon of information presuppositions we ought to recognize that the propositional context of utterance need not consist of shared assumptions at all. The speaker's presuppositions may be informative to the hearer; so in cases where the presuppositions are propositional elements of the context, those elements need not belong to the hearer's assumptions at all. This conclusion is another step in the direction of objective propositional contexts. [...] We may find, for instance, that a speaker's presuppositions are merely the speaker's own *take* on the context of the propositional context and the case of presuppositions informative to the hearer is that in which the hearer had not taken the objective prepositional context to contain what, as it turn out, the speaker takes it to contain.'

Our own analysis of Karttunen example can recognize all of the observations made by Gauker, without granting that those observations undermine the epistemological methodology at the root of the common-ground analysis, which we consider basically correct. Let's analyze an

(slightly modified) example proposed by Gauker in order to make this more explicit.

> Suppose I say to Alice, 'I will meet you in front of the department store at 6 pm'. If things go smoothly, then at 6pm Alice and I will both be standing near the entrance of Nordstrom's. In that case, we may say that Alice and I ended up at Nordstrom's because the content contained 'We will meet in front of Nordstrom's'. [...] But suppose things do not go smoothly, so that at 6pm I am standing in front of Nordstrom's while Alice is standing in front of Bloomingdale's. In that case, we may explain that there was a disparity between what the context really did contain and what one or the other of us took it to contain. In other words, one or the other of us had mistaken presuppositions. In such episodes of misunderstanding, the fault may lie with the hearer or, quite possibly, it may lie with the speaker.

The view of context in terms of shared assumptions says that the context of utterance is constituted by the shared presuppositions of speaker and audience. Gauker challenges this received view. For him contexts of utterance have an objective nature and presuppositions are not defined as elements of these objective contexts. An interlocutor's presuppositions are defined in terms of what (s)he supposes belong to the propositional context proper. So, according to Gauker's analysis, the utterance:

(U.) 'I will meet you in front of the department store at noon'

is supposed to be associated with a propositional context which can either contain 'The department store is Nordstrom's' or 'The department store is Bloomingdale's'. But there are numerous conceivable situations where there might not be any fact capable of objectively deciding which one of the sentences goes into the so-called propositional context. Nevertheless in these situations it is perfectly possible to envisage agents presupposing any of these sentences. Gauker's theory can still explain those cases in terms of a discrepancy between what the context did contain and what one or the other agents took it to contain. Nevertheless the theory seems unable to explain the nature of the agent's takes on objective context. The theory sketched above focuses on explaining this epistemic aspect of the problem - an aspect that need explanation even if one adopts a realistic stance. And the theory elaborates these takes in terms of what is taken to be commonly presupposed, rather than as takes on some objectively constituted aspects of context. Say that Christopher and Alice have planned to meet after work, have dinner and then go to the cinema. Say that Nordstrom's is

the department store closest to the agent's work places, that Bloomingdale's is the department store closest to the cinema they want to go, and that there is a 20 minutes commute between both stores. Say, in addition, that the better restaurants are around Nordstrom's, while around Bloomingdale's there are few restaurants of lower quality.

So, when Alice receives the information conveyed by (U) she might think that, although there is a small risk of ending up in a bad restaurant, Christopher might have presupposed that they will meet in front of Bloomingdale's. This inference involves assuming that Christopher is sufficiently risk averse (with respect to arriving late to the theatre) to prefer otherwise. But, of course, she might be wrong. Christopher might not be very risk averse in this case. Sufficiently so in order to presuppose that they will meet in front of Nordstrom's, widening the choice of good restaurants, while increasing the risk of arriving late to the theatre. And, of course, Christopher might underestimate the risk aversion of Alice. It is quite clear that in this case they do not coordinate.

So, in this case the speaker will associate (U) with the presupposition or 'The department store is Nordstrom's'. The hearer associates (U) with the presupposition 'The department store is Bloomingdale's'. Neither presupposition is a shared assumption because there are no shared assumptions about this matter. And it does not seem clear in virtue of what fact there should be some objective context containing any of those presuppositions. What is clear is that the agents fail to coordinate because they have false beliefs. But what they get wrong is not some presupposition that is supposed to be objectively associated with (U) in this exchange. What they get wrong (in this case) are the preferences of their interlocutors. One agent overestimates the risk aversion of its partner, while the other underestimates it.

The version of Soames' account presented above can nevertheless be used in order to model the previous scenario. Here the speaker is Christopher and Alice is the hearer. So, the proposition corresponding to 'The department store is Nordstrom's' is in *Pr*s because the speaker believes so and he thinks that this is uncontroversial. And he thinks so because he thinks that this proposition should be commonly presupposed. And, finally, he thinks so because he thinks that Alice should presuppose one of those two items, and he also thinks that when facing the decision of what to accept, she will choose Nordstrom's over Bloomingdale's. A similar analysis applies to Alice. Her presuppositions are determined in terms of her take on the content of the presuppositions

of the speaker.[21] We spare the reader the details of this analysis, since it proceeds a similar fashion, but with the opposite outcome.[22]

It should be said in passing that the body of shared assumptions plays an interesting hybrid role in conversational exchanges. I might presuppose at the moment that it is raining in Sag Harbor, and I might think that you take that for granted too, but I might be wrong about both facts. The body of our shared assumptions is something about which I can have false beliefs, as I can have false beliefs about the state of the weather in some location. There is, of course, an asymmetry between brute physical facts and bodies of common attitudes. My mental acts contribute to changes of those bodies of attitudes, while my mental acts have no causal relevance for changes in the weather. But the changes introduced in bodies of shared assumptions are *very* different from deliberate changes that I might decide to implement in my view. While we have some control over the latter changes, we often do not have complete control over the former. Only changes of shared attitudes implemented via highly regimented communication protocols are of the latter type, and those are infrequent in common conversation.

So, in common situations the participants in a conversation are uncertain about the contents of the body of shared assumptions and they can at most have beliefs about it. That this is the case seems to have been explicitly recognized by Beaver in a recent series of papers (see, for example, **?**). In this model what is updated by agents (following assertions) are beliefs about presuppositions shared in common. In terms of our model Beaver uses plausibility orderings in order to implement updates of P^i structures. Various common ground candidates are ordered in terms of how plausible it is that a particular common ground candidate is in fact the common ground assumed or intended by the speaker. Updates of this plausibility ordering with a proposition p select the most plausible p candidates. This is a particular mechanism for update first proposed by **?** and later on adopted by various computer scientists. Here we preferred not to adopt any particular mechanism for update, and not to say much about whether a qualitative procedure of the sort used by Beaver is sufficient to implement a variant of our model. Nevertheless, the reader can safely conclude form previous observations that a purely qualitative account of update will suffice only if it implements some suitable version of a qualitative theory of decision -

---

[21]In other words the hearer presupposes that p if she thinks that the speaker considers p as part of the record of in case the hearer thinks that the speaker thinks that the audience is prepared to add it, without objection, to the current context.

[22]Notice that the above considerations offer an account of *why* both speaker and hearer think that their respective presuppositions belong to the common ground.

incorporating not only a representation of belief, but also a qualitative version of preference or value.[23]

6.5 Implicature and the inferential encoding of doxastic commitments.

In the previous sections we adopted a skeptical stance towards the descriptive power of accounts of doxastic context closed under strong rationality conditions. The Stalnakerian account described at the beginning of this article requires the postulation of equilibrium states, which are supposed to be both logically closed and in introspective equilibrium (i.e. they are stable in the sense of being closed under the fixed-point equations S1-2). We argued that equilibrium sets of this sort can be employed in order to present a theory of what is normatively required of rational agents. In other words, stable sets of this sort are better conceived as representing the *doxastic commitments* of interactive agents. Rational agents should strive to reach equilibrium states of this sort, even when these states are rarely (if ever) instantiated in actual exchanges. The epistemic commitments of stable agents can be articulated in terms of an inference relation as follows:

(AC) B is an *autoepistemic consequence* of A if and only if B holds in every stable context for A.

In this section we will discuss an alternative interpretation of this notion of consequence offered by Robert Stalnaker in **?**. The gist of the proposal is:(P) what is autoepistemically entailed by a sentence α is what is meant or implicated, but not explicitly said by uttering α. He suggested that a defeasible notion of consequence (explicated via AC) could be used to formally encode Grice's notion of *conversational*

---

[23]Two observations need to be made here. Hinzen extracts in **?** more radical conclusions than the ones suggested here from similar observations about the fact that a full-fledged theory of decision seems to be required to identify contexts and determine their dynamics. The observations made here seem also valid even if the classical theory of propositions remains unchanged and one only focus on clarifications and extensions of the type proposed by Beaver.

The second point concerns the possibility of eliminating desires from the analysis by adopting a cognitive reduction, where desire is understood as a species of belief. A common variant of this strategy is to claim that the degree to which an agent desires any proposition p equals the degree to which the agent believes the proposition that p would be good. Nevertheless, David Lewis argued persuasively in **?** against such anti-Humean strategy. Further arguments of this type are offered in ACL (1995). Desires (cognitive values) seem to be needed in order to specify the dynamic aspects of context. And, when postulated, they seem hard to eliminate cognitively. Not at least without seriously disturbing the standard cognitive architecture presupposed by Bayesian epistemology.

implicature - and to understand its context-dependent behavior.

This final section makes three main contributions concerning this proposal. First, we will focus on the tenability of (P). Paul Grice considered in **?** some minimal constraints on implicature needed in order to handle G.E. Moore's paradox of 'saying and disbelieving'. We will show that (P) is incompatible with those constraints. Secondly we will offer an alternative account of AE consequence based on several remarks made by Grice in **?**. According to this account the AE-consequences of a sentence α encode the body of full beliefs to which someone is *committed* after uttering α. Thirdly we will offer a preliminary account of the formal properties of this new notion of consequence.

6.6 Autoepistemic logic: some background

First some clarifications about the underlying language. Let L_0 be a Boolean language and let L be the language formed inductively from L_0 by adding the formation rule:

$$\text{If A} \in L, \text{ then L(A)} \in L^{24}$$

L's intended interpretation will be left open for the moment - the reader can see it neutrally as one of the attitudes A postulated in section 1. The notation is reminiscent of the one used in modal logic for the necessity operator. This is basically the idea behind the operator although the emphasis in autoepistemic logic is, of course, epistemic rather than ontologically related.

The following notation will be useful later on (Γ here could be either a theory defined on L_0, L, or languages of intermediate complexity):

$$\text{L}\Gamma = \{\text{L(A)} : \text{A} \in \Gamma\}$$
$$\neg\text{L}\Gamma = \{\neg\text{L(A)} ; \text{A} \in L, \text{ and A} \notin \Gamma\}$$

An AE theory is any set of sentences in L. A crucial concept is the notion of *stable* set:

DEFINITION 1: A stable set Γ satisfies the following properties: (1) Γ is closed under logical consequence, (2) If A $\in \Gamma$, then L(A) $\in \Gamma$, (3) If A $\notin \Gamma$, then $\neg$L(A) $\in \Gamma$.

Stable sets are sometimes called AE theories. Robert Moore proposed (see **?**) the idea of a *stable expansion* of a premise set Γ. The gist of his proposal is to represent an agent whose epistemic state is

---

[24] In order to avoid misunderstandings between propositions and sentences, in the following sections we will use upper case letters, A, B, ..., etc, for sentences of the underlying language.

both stable and *grounded* in Γ. The basic idea is to characterize a set T containing at most the consequences of $\Gamma \cup LT \cup \neg LT$.

DEFINITION 2: A set T_Γ is a stable expansion of the premise set Γ if and only if it satisfies:

$$T_\Gamma = \{A : \Gamma \cup LT_\Gamma \cup \neg LT_\Gamma \vdash A\}$$

A set Γ_0 of non-modal sentences has exactly one AE extension, but modal sets might have various or no extension (for example, $\{L(A)\}$ lacks extensions and $\{L(A) \rightarrow A\}$ has two).

Now we have enough background to define a non-monotonic notion of AE consequence:

DEFINITION 3: $\Gamma \mathrel{\vdash\!\!\!\sim} A$ means A is contained in every stable expansion of Γ and there is at least one such stable expansion.

Of course, $\Gamma_0 \mathrel{\vdash\!\!\!\sim} A$ means that A is contained in *the* AE extension corresponding to Γ_0. Therefore, for modal-free sets of premises, $T_{\Gamma_0} = C(\Gamma_0)$. For modal Γ one needs to consider all possible epistemic contexts induced by Γ, i.e. all the AE extensions of Γ.

The previously defined notion of consequence does not enforce the inference from $L(A)$ to A. In other words the pattern:

$$(P2)\ L(A) \mathrel{\vdash\!\!\!\sim} A\text{[25]}$$

does not hold. Nevertheless the pattern:

$$(P1)\ A \mathrel{\vdash\!\!\!\sim} L(A)$$

is indeed enforced, because every autoepistemic extension of $\{A\}$ contains $L(A)$. The rationale behind this asymmetry is that most autoepistemic theories intend to model a notion of 'weak belief', rather than a notion of certainty or 'full belief'. [26] Rational agents should be self-aware of the facts they (firmly) hold as true. Nevertheless agents might believe facts that they are not willing to hold as true.

---

[25] If the existence of at least one relevant expansion is not required in DEFINITION 3, (P2) is actually enforced. Although the proviso is not commonly required in the literature we added it here. This seems to be required by a charitable reading of the literature.

[26] Autoepistemic logicians have not paid much attention to similarly motivated notions of consequence where the L-operator is informally interpreted as a notion of certainty, full belief or holding true. This is an unfortunate situation taking into account the centrality of this notion in decision theory and pure epistemology (as well as metaphysics). Preliminary considerations concerning the properties of the autoepistemic closure of sets of full beliefs are presented below. See also **?**, **?**, **?** and page 12 of **?**.

Stronger notions of consequence have been also considered in the literature. For example Kurt Konolige proposed the following alternative:

DEFINITION 4: $\alpha \Rightarrow \beta$ iff whenever α is in an AE extension, so is β.

This notion is certainly stronger. As a matter of fact it is monotonic. For assume that A $\Rightarrow$ B. Then assume by contradiction that A $\wedge$ C $\not\Rightarrow$ B. Then there is an AE extension E such that A $\wedge$ C $\in$ E and B $\notin$ E. But if this is so A $\not\Rightarrow$ B. For there is an extension (E) containing A that does not contain B.

6.7   Implicature and autoepistemic inference

The inference from A to L(A) is a very robust pattern enforced by all the notions of consequence considered above (and some alternatives considered below). In fact,

$$(\text{P1}) \text{ A } \mathrel{|\!\sim} \text{ L(A) and (P1') A} \Rightarrow \text{L(A)}$$

The intended epistemological interpretation of these patterns was discussed at the end of the previous section (the idea being that agents should be self-aware of the facts they (firmly) hold as true).

It is less clear how these patterns can be interpreted when we take into account Stalnaker's pragmatic reading of $\mathrel{|\!\sim}$ as a notion of (generalized) implicature. One might perhaps try to accommodate (P1) by arguing that part of the implicit speaker's meaning of every utterance A is determined by the speaker's belief in A. After all, if the utterance of A is sincere (and we can restrict our attention to this case) the main *reason* for the utterance is the speaker's belief in A. Nevertheless the grounds on which a sincere utterance is performed should not be confused with what is entailed or implicated by the utterance in question. Implicatures are calculated in terms of the suppositions needed in order to maintain the assumption that the so-called Cooperative Principle is observed (see essay 2 in **?**).

At this juncture it seems appropriate to quote Grice extensively about this point. His 'further notes' on his seminal article *Logic and Conversation* are highly relevant for the issue at hand:

> When I speak of the assumptions required in order to maintain the supposition that the Cooperative Principle and maxims are being observed on a given occasion, I am thinking of assumptions that are non-trivially required; I do not intend to include, for example, an assumption to the effect that some particular maxim is being observed. This seemingly natural restriction has an interesting consequence with regard to Moore's 'paradox'. *On my account, it will not be true that*

when I say that p, I conversationally implicate that I believe that p; for to suppose that I believe that p (or rather think of myself as believing that p) is just to suppose that I am observing the first maxim of Quality on this occasion. I think that this consequence is intuitively acceptable; *it is not a natural use of language to describe one who has said that p as having, for example, 'indicated', 'suggested' or 'implied' that he believes that p, the natural thing to say is that he has expressed the belief that p. He has of course committed himself, in a certain way, to its being the case that he believes that p*, it is bound up, in a special way, with saying that p.[Italics are mine, **?**, pages 41-42.]

The moral of the passage seems to be that Grice strongly opposes any possible formalization of his notion of implicature capable of sanctioning (P1). Therefore an important epistemological obstacle against reading AE-inference as a notion of implicature is that such reading does induce (P1).

The problem reappears if we consider some of the features needed for an implicature to qualify as a *conversational implicature*. One of these features is *cancelabilty*.

[...] a putative conversational implicature that *P*, is explicitly cancelable if to the form of words of the utterance of which putatively implicates that *P*, it is admissible to add, *but not P*, [...]

So, if we focus on *generalized implicature*, anyone who says *Pete is meeting a woman tonight* normally implicates that Pete is meeting someone other than Pete's spouse, sister or mother (Grice adds close Platonic friends to the list). The fact that this putative implicature does indeed qualify as a bona fide implicature is corroborated by the fact that it is perfectly proper to say:

Pete is meeting a woman tonight, but the woman is none other than his wife.

Notice that it is also perfectly possible for Pete himself to cancel the generalized implicature carried by the form of words: 'I am meeting a woman tonight'. In fact, Pete can felicitously say:

I am meeting a woman tonight, who does not happen to be anyone other than my own wife.

Nevertheless it seems that accepting (P1) as a constraint on a formalization of implicature will force us to say that the implicatures sanctioned by (P1) are cancelable by speakers only on pain of incurring G. E. Moore's paradox (see **?**). In fact, say that X says 'It is raining'. If the putative implicature is that X believes that it is raining, then it

should not be felicitous for X to say:

(M) It is raining, but I do not believe it.

Which is a form of Moore's paradox (usually called a paradox of 'saying and disbelieving').

The problem seems difficult to avoid because (P1) is a very robust AE inference preserved under different variants of AE logic. In particular the inference is granted by the type of AE theory envisaged by Stalnaker in **?**.[27] One possible way out could be to deny Grice's own thoughts about implicature and epistemic paradox, but this way out does not seem feasible. In fact, Grice's arguments are quite persuasive and they seem corroborated by other considerations (like the argument in terms of cancelability offered above). Another possibility could be to remark that (M) requires an indexical use of the belief operator that might not be encoded in the autoepistemic operator L. This way of circumventing the problem does not seem available either. This is so for several reasons. Perhaps the more obvious line of response can be based on the fact that several authors have suggested that AE-operators can indeed be interpreted in a multi agent context as indexical doxastic operators. We will not analyze nevertheless this issue in detail here (see Arló-Costa (1999) for a preliminary consideration on this problem).

There is also a second obvious line of response for the indexical problem. Even if one does not interpret L indexically, the argument in terms of cancelability can be run for the speaker. Perhaps it is a little bit odd for a third person to say: 'It is raining, and Pete just sincerely and accurately said so, but nevertheless he does not believe it'. Nevertheless this is perhaps not fully paradoxical. But when it comes to Pete himself, he can only cancel the putative AE-implicature by uttering (M). And this is a fully paradoxical.

The interest of Grice's insights is that they also give us an idea of how to understand autoepistemic entailment. In fact, Grice remarks that anyone uttering p also has 'committed himself, in a certain way, to its being the case that he believes that p.' In other words, it is perhaps plausible to say that what follows (classically) from the AE-extension of (a purely Boolean) premise set Γ_0 formalizes the *epistemic*

---

[27]**?** sketches a theory according to which the epistemic context induced by A (in a non-monotonic inference from A to B) is determined by 'all that the agent knows' at certain instant. The L-operator is, in turn, interpreted as a third-person operator of (weak)-belief. Without entering into the details of the proposal, it is intuitively clear that, as long as one endorses the idea that knowledge entails belief, the inference from A to L(A) should be made valid in Stalnaker's framework.

commitments contracted by accepting Γ_0.[28] This is perfectly in line with our own interpretation of the usage of context sets as encoding *doxastic commitments*. Autoepistemic inference seems to fall short of formalizing the more complex notion of implicature. But it can help to clarify the theoretical role of context sets.

This interpretation of AE-inference can make sense of the problematic inference pattern (P1). We can paraphrase Grice and say that any rational agent who accepts A should 'commit himself to its being the case that he believes that A'. And this commitment is the reason stopping him from canceling a putative implicature induced by A. Since the agent is committed to L(A), denying L(A) will put him in an incoherent scenario.

6.7.1 Certainty and belief

In the previous section we sketched an alternative to Stalnaker's account of AE inference. The idea (inspired by Grice's remarks on commitment) is that the AE consequences of a premise set A make explicit the doxastic commitments contracted by agents holding A true (or accepting the premise set A as given). We also considered above some possible objections to this interpretation. In this section we will focus on some residual problems importantly related to open foundational problems in AE logic.

It is easy to see that:

(A) $\emptyset \hspace{0.2em}\mid\hspace{-0.5em}\sim$ (T), where (T) is the alethic axiom L(A) $\rightarrow$ A

This follows from the fact that every AE extension of a set A is a stable set containing A and that every stable set is a S5 theory. Should we then say that every rational agent is committed to the 'alethic' axiom (T)? Remember that so far we are simultaneously maintaining that the L-operator encodes a notion of weak belief (instead of a notion of certainty or full belief). A negative answer to the former question is part of the philosophical folklore.

Kurt Konolige puts the problem in the following terms:

> AE logic assumes that agents are ideal introspective reasoners who do not subscribe to the principle T. [...] It is troubling, however, that the schema T is satisfied *post hoc* in any extension, but cannot be used in the reasoning to arrive at the extension. In metatheoretic terms, AE logic is not *cumulative* (**?**, pp. 228-229).

A cumulative logic is one obeying the principle also called *cautious monotonicity* stating that A, B $\hspace{0.2em}\mid\hspace{-0.5em}\sim$ C holds whenever A $\hspace{0.2em}\mid\hspace{-0.5em}\sim$ B and A

---

[28]Similar ideas have been defended by other authors in different contexts. For example see **?** or **?**.

$\vdash\!\sim$ C hold. But it is clear that cautious monotonicity fails when A is $\emptyset$, α is a non-tautological formula, C is $\neg$ L(α) and B is the following instance of schema T: L(α) $\to$ α.[29] To be sure there might be good reasons to construct inductive machines which are not cumulative, but the failure of cumulativity in AE logic seems to appear rather artificially as the result of forcing a particular doxastic interpretation on the L-operator used to build AE extensions. This interpretation is forced on L in spite of basic logical aspects of stable theories (the fact that they are S5 theories). Autoepistemic logicians have managed to implement smart moves in order to circumvent this mismatch of intuitions, but, as Konolige makes clear in his remark, the problem persists. The problem also is an epistemological obstacle against our 'commitment view' about the nature of AE inference.

In this article we will consider the following solution. We will define an alternative notion of AE inference, which obeys cumulativity and fits our 'commitment account' of AE inference. In addition we will propose a different pre-systematic understanding of 'L' as an operator of full belief or certainty. Pros and cons of this approach will be discussed below. But first we need some basic definitions used in AE logic.

DEFINITION 5: A stable set S is *minimal for a premise set A* if S contains A and there is no other stable set S' containing A such that S' $\cap L_0 \subset$ S $\cap L_0$.

Minimal stable sets have been considered in the recent literature in AE logic (see for example **?** pp. 234-5). To be a minimal stable set for A (MSS for A) is a necessary condition for a set of modal sentences to qualify as the set of introspective epistemic commitments associated with A. So a natural application of the ECP in this situation will lead to the construction of a notion of inference such that B follows defeasibly from A whenever B is in all the MSSs for A.

DEFINITION 6: $\Gamma \approx$ A means A is contained in every minimal stable set for Γ and there is at least a stable S such that $\Gamma \subseteq$ S.

This notion is obviously non-monotonic (consider the case when Γ is empty A is $\neg$ L(A), A is a propositional atom not included in ST(), and Γ is augmented with A). This notion of consequence obeys cautions monotony. To establish this fact we need some previous definitions.

---

[29]To see that the conclusion of cautious monotony fails it is useful to keep in mind that $\{L(\alpha) \to \alpha\}$ has two extensions: one including α and another which does not contain α. Therefore it is not true that $\neg$ L(α) is in every AE extension of $\{L(\alpha) \to \alpha\}$.

Let me first consider a class of *normal form* formulae. Consider the language L_D such that non-modal (purely Boolean) formulae are in L_D, if $\alpha \in L_D$ and $\beta \in L_D$, then $\alpha \vee \beta \in L_D$, and both $L(\alpha)$ and $\neg\, L(\alpha)$ are in L_D, when α is Boolean. We are basically focusing on formulae of the form:

$$\neg\, L(\alpha) \vee L(\beta_1) \vee L(\beta_2) \vee ... \vee L(\beta_n) \vee w$$

where α, β_i and w are all non-modal sentences.

It can be established as a lemma that every consistent and stable A-theory K (where $A \in L_D$) is such that there is a minimal stable set for K. This lemma can, in turn be generalized with the help of the fact (proved in **?**, page 230) establishing that every set of modal sentences (containing the L-operator) is K45-equivalent to one in normal form.

Consider now Cautious Monotony. Assume that both A $\mathrel{\vmid\sim}$ B and A $\mathrel{\vmid\sim}$ C - where $A \in L_D$. We have to check that $C \in K$ where K is a MSS for A $\wedge$ B. Notice that (in the presence of the assumptions) every MSS for A $\wedge$ B is also a MSS for A. For pick an arbitrary MSS for A $\wedge$ B, say K, and assume by contradiction that it is not a MSS for A. Then (since K is an A-theory) there is a stable set K' such that $A \in K'$ and $K' \cap L_0 \subset K \cap L_0$. Now, in virtue of the previous lemma, either K' is itself a minimal A-set or there is K'' (such that $A \in K''$ and K'' $\cap L_0 \subset K' \cap L_0$). Since A $\mathrel{\vmid\sim}$ B, $B \in K'$ (or K''), which contradicts the assumption that K is a MSS for A $\wedge$ B. Now, since K is a MSS for A, and we assumed that A $\mathrel{\vmid\sim}$ C, $C \in K$, which suffices to complete the proof. Almost identical strategy can be applied to establish that $\mathrel{\vmid\sim}$ obeys CUT and other basic properties of non-monotonic inference.[30]

Of course, $\mathrel{\vmid\sim}$ obeys the pattern (P1):

$$(\text{P1}) \; A \mathrel{\vmid\sim} L(A)$$

It is also the case that $\mathrel{\vmid\sim}$ obeys:

$$(\text{P2}) \; L(A) \mathrel{\vmid\sim} A$$

This second pattern, according to our understanding of $\mathrel{\vmid\sim}$, indicates that every agent who accepts $L(A)$ is committed to hold A true. We understand here (tacitly) that the occurrence of non-modal sentences in a stable set represents the fact that the agent in question (firmly) holds (or takes as) true the sentence in question. On the other hand we are leaving open which is the intended meaning of the L-operator compatible with our characterization of $\mathrel{\vmid\sim}$. Nevertheless it is not difficult to elicit this meaning taking (P2) into account. L cannot be understood

---

[30]Cut establishes that A $\mathrel{\vmid\sim}$ C follows from A $\mathrel{\vmid\sim}$ B and A, B $\mathrel{\vmid\sim}$ C.

as a notion of weak belief. It should be understood as the strongest doxastic attitude of which the agent is capable. Basically L in this context should be understood as a notion of certainty or full belief.

In this context the troublesome inference $\emptyset \hspace{2pt}\mid\hspace{-4pt}\sim$ (T), ceases to be problematic. In fact, such a pattern only says (in this context) that every agent who fully believes A is committed to the truth of A. It is important to realize that the pattern $\emptyset \hspace{2pt}\mid\hspace{-4pt}\sim$ (T), should not be interpreted as saying that AE agents are committed to a principle of 'arrogance'[31] legislating that full beliefs are true. What the schema T says in this context is that agents are committed to *accept* that fully believed items are true. In other words, form the point of view of the agent, every item of which he is certain is an item such that he should be willing to (firmly) hold as true. Of course, the item might be false (objectively), but all that counts for AE inference are the introspective obligations of the agent in question (at a certain instant). Therefore rational agents should accept schema T as a coherence principle in charge of maintaining an equilibrium between the items held as true (non-Boolean sentences in his stable theory) and his certainties (sentences prefixed by L-operators in stable sets).

Of course the inference (T) is problematic if L is not pre-systematically understood as a notion of full belief. Why a rational agent should be committed to (firmly) hold as true items which are only (weakly) believed? In contrast it seems reasonable to require that rational agents should be committed to hold as true every item of which (s)he is certain of (or which (s)he fully believes).

AE logicians seemed to have paid attention only to either a variety of notions of weak belief or to the strong notion of knowledge, somewhat neglecting the intermediary notion of full belief. Such a notion is widely used in many fields where doxastic representation matters (the theory of games and decisions is an obvious example of a field where the notion plays a crucial role). This section has been devoted to define a cumulative notion of AE inference which seems optimally understood when the intended interpretation of the L operator is as an operator of full belief. The interpretation has the independent virtue of offering an unified explanation of the role of schema T. The schema is indeed satisfied in every stable theory (as a matter of fact in every AE extension) and the principle can be cumulatively used in AE reasoning.

In addition the picture fits Grice's account of pragmatic inference in terms of commitments, and it circumvents the previous criticisms to the idea of seeing non-monotonic inference as the encoding of pragmatic

---

[31]The terminology is used in **?**.

implicatures. Finally, our definition of $\mathrel{\mid\approx}$ satisfies several of the widely accepted properties of non-monotonic inference (like cumulativity) in a rather natural manner.

Acknowledgements: I would like to thank Andrea Bonomi for very insightful comments and Mandy Simons for various useful pointers to relevant literature in Linguistics related to context.

References

1995. *Desire-as-Belief Implies Opinionation or Indifference*.

Arló-Costa, H. 1999. Qualitative and probabilistic models of full belief.

Arló-Costa, H. forthcoming. First order extensions of classical systems of modal logic: the role of the barcan schemas. *Studia Logica* .

Bovens, Luc and Stephan Hartmann. 2001. Belief expansion, contextual fit and the reliability of information sources. In V. Akman, P. Bouquet, R. Thomason, and R. A. Young, eds., *Modeling and Using Context: Third International and Interdisciplinary Conference, Context 2001*, pages 425–428. Berlin: Springer-Verlag.

7

Context and Logical Form

Wolfram Hinzen

7.1 Introduction

Contextuality is a pervasive phenomenon in the study of cognition. Nonetheless, the way in which laws for the justification of induction, say, are context-dependent, might not relate in any direct manner to the way in which, say, meaning is contextual. Thus we won't be able to presuppose a general notion of context, or of the "problem of context". We will have to ask for each aspect of cognition separately what the notion and the role of context is.

This chapter analyzes contextuality as relevant to a science of meaning. I will first, in section 2, introduce a notion of meaning deprived of contextuality altogether. I then discuss some aspects of "dynamic" semantic theories, for which the contextuality of meaning is decisive (section 3), and argue that these models naturally lead to normative models of discourse, considered as a kind of human action. In section 4 I argue against Stalnaker's recent criticism of dynamic semantics and the representation of context he himself recommends. Finally, in section 5, I explore in what sense there might be a *linguistic* (sentence-internal) notion of context that is important to linguistic explanation.

7.2 Meaning as Non-Contextual

7.2.1 Meaning

How might meaning be contextual? I will take this to be a question about *linguistic* meaning, a notion that describes aspects of the interpretation of natural language expressions as determined by the rules of

language themselves. Note it is thus, in particular, not a notion applying to features of *utterances*. By an *expression* I will mean a pairing of a sound and a meaning, and by a *language* a systematic way of generating such pairings.

The existence of a notion of linguistic meaning thus understood might be denied, for example on the ground that the rules of a language never determine any aspects of interpretation, or that they do it only together with non-linguistic factors. On such views, an interpretation accrues to a natural language sentence at least in part *externally*, for example by virtue of the fact that on an occasion of its use, a speaker of that language expresses certain beliefs or opinions with it. But there would be no interpretive aspects of expressions that are linguistic narrowly speaking and that we could isolate as such.

But if linguistic apects of meaning cannot be isolated from various non-linguistic ones that also determine aspects of meaning on an occasion of use, meaning is a rather messy phenomenon. How a speaker on an occasion of the use of an expression ascribes a certain interpretation to it depends on a myriad of factors and cognitive faculties. Scientific progress depends on suitable idealizations and theoretical terms that make lawful generalizations statable. These will usually depend on decontextualizing a phenomenon as complex as "interpreting an utterance in context".

If, on the other hand, the claim is not only that linguistic aspects of interpretation cannot be isolated from non-linguistic ones, but that no aspects of interpretation *are* intrinsically linguistic, it is clearly false. For that in example (1)

(1) [John expected to hit himself],

himself is necessarily (by the rules of language) the same as *John*, is clearly an interpretive property of (1). It follows from general aspects of its grammatical structure, in which *John* binds the reflexive. It is equally intrinsic to

(2) I wonder who [John expected to hit himself],

that in it *himself* necessarily isn't the same as *John*. That again follows from structural properties of it, namely that the reflexive is bound by a variable, which in turn is bound by *who*: a paraphrase of (2) might be

(2P) I wonder for which x, x a person, John expected x to hit x.

Note that these examples do not depend on a solution to the notoriously difficult question to free a notion of linguistic meaning from other ingredients of interpretation such as background beliefs, world knowl-

edge, and other non-linguistic factors. I have no definition of "language" on offer here, except that I stipulate that underlying language use there is a system of grammatical knowledge represented in the mind. It seems clear enough from examples and actual work what a grammatical process or a rule of language is, and to the extent that the examples show such processes or rules to have interpretive effects, they give content to the notion of linguistic meaning.

It seems hard to question then, not only that there are interpretive aspects of expressions which are linguistic, but also that they are *non-contextual*. For the relevant interpretive aspects of (1) and (2) have nothing to do with the non-linguistic context in which the expressions are used. They are things that speakers of English *know*, being consequences of an abstract system of principles and rules. To the extent that the explanation of standard facts about analyticity, synonymy, or ambiguity that are known to a native speaker proceeds on the basis of such principles and rules, it would seem that non-linguistic context drops out from the theory of meaning. Whether a sentence is ambiguous, for example, would, if so explained, have nothing at all to do with how or why it is used, and how contextual the interpretations assigned to such uses may be. It would be explained by appeal to the fixed structure of an internally represented system of human knowledge.

This then is the question we ask in this chapter: to what extent there *are* interpretive aspects of expressions which are linguistic but which *do* depend on the non-linguistic (epistemic, environmental, etc.) contexts in which they get used. I will discuss some recent views on which content and context become a unified phenomenon, and argue that on this view, meaning, far from being a linguistic notion, merges into a general notion of "speech act content". The latter seems far too contextual to be theoretically tractable, on the other hand, shot through as it is by pragmatic inferential mechanisms for fixing it. Arising in part through general principles on how to act in a discourse context, it is open to normative description, much like human action in general. As such, it is not a notion on which we can build linguistics as a (naturalistic) science.

7.2.2 Logical Forms (LFs)

I will assume, in the tradition of generative grammar, that in a speaker's mind/brain the interpretive properties of expressions are represented in a component of the language system that is standardly called *Logical Form* (LF). It is by appeal to the structure of the representation of an expression at that level of representation LF that we explain certain linguistic aspects of its meaning. To take a standard example, the

sentence

(3) Pictures of everybody are on sale

can mean that everybody is such that pictures of him/her are on sale, or that pictures are on sale that have everybody on them (group pictures). But importantly, (3) cannot mean: for some pictures, x, everybody is such that x is a picture of him/her and x is on sale. How do we explain these empirical facts? In logic, we make the difference merely manifest, by associating two different logical forms with the expression, with appropriately different relative scope relations among its quantifiers. This is a *representation* of the relevant fact, not an explanation (and it doesn't tell why the third reading is not possible). If we add that these different logical forms are associated 'by convention' to the expression (3), we make no further step towards explanation: there are no theories of convention that tell us how this works, and it seems imlausible in any case that highly constrained syntactic principles responsible in human languages for what scopes over what are the result of arbitrary conventions.

Generative grammar takes a different course, and attempts to *explain* the different logical forms associated with an expression by deriving them by syntactic means - mechanical processes like merging two expressions, targeting one syntactic structure so as take one constituent from it and insert it in a different place, and so on. It is by means of such processes that the quantified noun phrases *pictures* and *everybody* in (3) get their relative scope possibilities. The impossibility of the third reading is explained by the fact that its derivation violates independently motivated constraints on possible linguistic structures.[1]

Note then that the notion of LF in its linguistic sense has little to do with logic, normative rules of inference, or commitments we undertake: each LF is a natural object in a speaker's mind/brain, and that it plays a role in logic and reasoning, if it plays such a role, is not what determines its properties. [2] It is important to note that the claim that there is such a thing as a component of LF in the human language system is an empirical claim that is right or wrong. By contrast, there is a sense in which a logical form in the sense of logic or the theory of reasoning comes into existence through its definition. It is clearly not a notion of natural language meaning, being a characterization of *thought*, which presumably is what Frege and Russell designed it for. [3]

---

[1] See Huang (1995), 132-135 for details.

[2] At best, it provides evidence for the linguist to ascribe certain properties to it.

[3] Normatively construed commitments as in Arló Costa's contribution (this volume) can have no role to play in a strongly naturalistic enterprise like the present

For these reasons I will adopt a more recent label and speak not of logical forms, but of a component of *Semantic Representation* in the human faculty of language, SEM for short. To say then that there are interpretive aspects of expressions which are contextual would be to say that there are features of the extra-linguistic context that make linguistic expressions have certain properties at SEM. Considering its conception in recent grammatical theory (Chomsky (1995)), there is a sense in which SEMs are not open to doubt. It would be the claim that grammatical operations make no contribution to semantic interpretation, language being pure sound. As argued above, this option seems mistaken. What might well be wrong is that SEM is the only interface at which the language system meets other cognitive components governing performance. That however is an empirical question, however hard to answer, and for the moment the very best assumption one can make is the minimal one, according to which there is only one, and try to prove this assumption wrong. If so, SEM will not be negotiable at least for the moment.

I will suppose further that human languages consist of a list of primitive or unstructured *lexical items* (LIs) as distinct from a computational system (CS). A (mental) lexicon, on which the rules of the latter operate is a storage place for the sound and meaning properties of lexical items. Let these properties be expressed in terms of a collection of features. Each LI is a collection of *features* of different sorts, a representation of what a speaker knows about them, and it is these features which are fed into the computational component of language, which builds larger representations out of them.

These assumptions make the above contextual option rather inplausible. Derivations in the faculty of language are encapsulated and have locality properties, first in the trivial sense that what the computations have access to are symbolic structures represented in the mind alone, and secondly in the sense of the empirical assumption that they proceed in the mind's linguistic component alone. On its way to compute a SEM, the CS doesn't have access to what is happening outside the head, or elsewhere in the head. What enters into a derivation is mental representations of a specific format that CS can compute, but non-linguistic

---

one. The distinction between old Chomskyan 'competence' and Arló Costa's 'commitment', is crucial here, although both are contrasted with 'performance'. It will transpire further below, especially in the last part of this paper, that Arló Costa's proposed marriage between theories of linguistic meaning and normative models in formal epistemology should be made consistent with the fact that all aspects of meaning depending on the generative structure of expressions have nothing to do with the epistemological states of speakers.

context doesn't enter the computation of SEMs at all. This is more or less also the conclusion that I will reach below, even though there remain some interesting issues concerning the role of context variables *within* SEMs.

Jason Stanley has recently stated that "all effects of extra-linguistic context on the truth-conditions of an assertion are traceable to logical form" (Stanley (2000), 396). The sense of this claim will depend on clarifying, of course, the notion of logical form used. Stanley explains that it is the "'real' structure" of a sentence, which may be "quite distinct from its surface grammatical form" (392). That leaves things rather open. In its generative-grammatical sense, in particular, the notion of LF is hardly describable in Stanley's terms, for whom it seems to depend on notions of Deep and Surface Structure long abandoned. The Minimalist Program (MP, Chomsky (1995)) at least is also not coherent with Stanley's assumption that LF is a structural representation that is the object of compositional and referential semantic interpretation (cf. Stanley (2000), 399). While the notion of truth conditions is still informally used in MP, it is not *essential* to the theory of LFs that they determine such truth conditions. LFs are internally motivated, by the operations of a grammar. Truth conditions provide *data* for a theory of LF, but they are things to be *derived* on syntactic grounds (cf. the above example). Moreover, if informally used, the notion of truth conditions is, contrary to what Stanley assumes (e.g., 395), applied to *expressions*, which are objects, and not to assertions or utterances, which are acts.

Logical form being a highly theoretical term fixed by its definition in technical discourse, Stanley's main question will at least in part have an answer that simply follows from this definition. Thus if Stanley suggests that "all effects of extra-linguistic context should be traced to logical form", we might take him to suggest as theoretically useful a specific theoretical notion, rather than to answer a factual question. Contemplating its usefulness, it seems Stanley's notion is very rich in terms of the information it captures.

Nonetheless, there is some affinity in what he says to what I will myself claim in section 5. This is the idea that "the effects of context on the truth-conditional interpretation of an assertion are restricted to assigning the values to elements in the expression uttered" (396). If only this idea didn't depend on the logical notions of "truth condition", "semantic value" and "denotation" that seem questionable as notions of linguistic theory (on which more below), it seems congenial to the proposals to follow. That is to say: the role of context is the instantiation of certain variables occuring in SEM, and it doesn't contribute to linguistic meaning in any other way. I will now delve into

some of the ways in which the effects of extra-linguistic context on the interpretation of human speech is vast.

7.2.3 SEMs and Context

Context enters when SEMs are put to use by performance systems accessing the language system. The moment where language gets produced and interpreted, contextual effects abound. In a context the system of linguistic knowledge interfaces with human intentions, background assumptions, culture and custom, a perceptual situation, the climate, human temper, and more. Maybe someone using the word *sun* in the Sahara will use it with slightly different connotations than when he uses it in England. Maybe a great composer will say something slightly different from what we say when he says

(4) I like music.

Maybe he will *like* it in a different way than we do, and think of something quite different when he uses the word *music*, even though that might depend on whether he does it on a good or a bad day. All of this is possible, but if all of it is meant to be part of the *meaning* of what he says, that notion would appear to be rather inscrutable. We might put together the "meaning" of the word *music*, in this massively contextual sense, out of what he means and we mean and your friends mean and maybe what other populations mean, at certain times or others. Whatever we fabricate here would not seem to result in anything of theoretical interest.

Neither do the examples just put forward show anything about the contextuality of linguistic meaning. All of the contextuality just observed is perfectly consistent with the view that concepts are innate (and thereby non-contextual). [4] On one such view, humans are born with a stock of concepts that get phoneticized upon exposure to an environment. Human nature apparently makes them similar enough to make communication reasonably effective, but that doesn't show anything about how much variance we may find across speakers and contexts, which may have different sources. The above observations could be redescribed thus: Even though, by way of innate endowment, the concept *music* is whatever it is, it just happens that contexts of language use embed the knowledge of *music* as specified in the mental lexicon of a speaker A in a rich background of opinions and feelings that A has about music in those contexts. Not A's concepts vary, but his feelings. The same happens to speaker B. Some of the opinions and feelings of A and B may be shared, but that doesn't mean that beyond

---

[4]E.g. Fodor (1998) for one such view.

what's in the head of A and B there is some mysterious shared concept of music outside the heads of both (I return to this below).

Maybe in fact we have slightly different concepts in different contexts, which we associate to the same sound. But maybe not. The feature of human languages that Jerry Fodor calls their *systematicity* seems to depend on the fact that a word in one context means what it does in another. It's because music means the same in *I like music* and *I don't like music* that both sentences cannot be true together. Relative to an idiolect a word will have the interpretation it has, and it is with this interpretation that it enters syntactic compounds, to which it contributes precisely its interpretation, and nothing else. In each context of use it may happen that music is identified or valued in different ways. But what matters for lexical theory is the content a word contributes to the content of the compound in which it enters, and it doesn't seem to be the case that epistemic features of a context, changing criteria of identification, and so on, enter systematically in the formation of syntactic compounds, in which they seem invisible.

Thus it is everything but clear whether even upon adding a lexicon to a syntactic theory we open up our investigation to an intrusion of context that might be fatal to a naturalistic scientific endeavour regarding human language. That conclusion clearly depends on a choice of subject matter for linguistic theory, namely the idiolects of speakers (or I-languages, in Chomsky's term). For syntactic theory, that choice is perfectly standard, and I can't see how it could be negotiable for the theory of the lexicon. What would the units of linguistic computation be, on a more communication-theoretic view, say? Where the theory of the lexicon is part of a theory of I-language, concerned with the unstructured units that get computed by the computational system of human language, the lexical concepts are characterized by features, which are syntactically non-structured mental representations. But on the communication-theoretic view, how do we fix a concept? The relevant signals here - phonetic stimuli - will typically differ from speaker to speaker, and from dialect to dialect. Symbols which humans use to convey thoughts are rather abstract, contrary to animal communication - say the dogs' wagging of tails and their leaving scents on trees - where we have clear cues of a salient optic and chemical sort. The latter cases make it seem plausible that, eventually, signals can be matched with interpretations; in the case of the symbols that humans use this is less clear. [5]

---

[5]To my knowledge there is little by way of theories of convention that would explain how interpretations are attached to signals. Even the simplest human concepts are of an extraordinary complexity, raising poverty of stimulus problems of

7.2.4 The Mystery of a Shared Language

The facts noted with regard to differences in phonetic stimuli and interindividual differences in conceptual understanding might provoke the thought that there *must* be some shared language in the sense of a "public code" which would relate "public pronounciations" and "public meanings", or else we couldn't communicate. The elements of this code will have properties that we all for some reason agree on. This is a rather abstract construct, more abstract than the I-languages mentioned above, which they presuppose and which are concrete structures internal to an organism.

But take the word *disinterested*, which most speakers use to speak of someone who is uninterested, although a few others take it to mean "unbiased", or "unselfish". In this concrete case, we somehow communicate in using the word, but the codes are different. Somehow we seem to be fortunate enough that human communication is mostly as efficient as dog communication. Nonetheless, who or what decides over what code one uses? One old answer appeals to "comunicative intentions", so that what I *intend* or *decide* to convey is what fixes what my words mean. But by bringing in the notion of intention - close as it is to the notion of meaning - we provoke more puzzles rather than less.

Efficient communication doesn't seem to *depend* on an exact matching of interpretations. People partially converge in assigning interpretations to signals, but on the whole communication and understanding is a matter of more or less. As in the question of what it is to study the notion of "being near", it seems unclear what it is to study "communication", if this is to be a scientific account, not a matter of history or human affairs. Maybe the more assumptions Mary and Joe share the better they get along in using their respective generative linguistic faculties, and maybe it happens that the more they talk, the more they will converge on background assumptions. But it seems wrong-headed to stipulate that at some early point of such an interchange the people "do not (yet) communicate", that is, that there is a minimal amount of a shared background that *has* to be there for a linguistic exchange to deserve its name. Whether it does or does not deserve that name seems like a (contextual) matter of human evaluation. If so, language is not, by its essence, social. Nothing *has* to be shared (although nobody denies of course that language use is a social phenomenon). A speaker

---

the strongest form, considering also the rate of concept acquisition (see Chomsky (2000), 63-66 for relevant comments). This suggests that concept acquisition is merely a labeling of concepts pre-existing in the mind. Labeling in this sense would be a brute-causal process, in which matters of rationality and conventional agreement play as little a role as in matters of the acquisition of syntax.

who says

(5) My dog is happy today

may be referring to a dog he owns, a dog he is making a walk with, a dog he is seeking to purchase, a dog he sketches on a canvas, or something else. Whether what he refers is "in fact" a dog, and whether it is *his*, is a matter of human value, much as is a judgement about whether the utterances was "appropriate", given, say, that it was about a dog that is being painted on a canvas. Interpreters may want to figure out what exactly *my dog means* here, maybe with uniform results, maybe not. It is not some "shared expression" *my dog* (in a public code) which "generates a presupposition" (a phrase of Robert Stalnaker's, to whom I will return). It is practical human reasoning that will decide, on an occasion of use and in ways that are independent of specifically linguistic principles, whether a speaker of (5) is taken to *own* a dog. Similar remarks apply to *happy* and *today*.

Thus we derive the preliminary conclusion that there is a non-contextual study of meaning in the form of a study of the language system's SEM, but that the contextual aspects of the meanings of public speech acts - although no doubt they exist, as a matter of historical narrative, of norms and values - do not seem to lend themselves to a sort of scientific analysis that would be continuous with the study of SEM.

7.3   Content as Context

7.3.1   Dynamic Semantics

Semantics, the formal study of meaning, has mostly made language-world relations its topic, but has also tried to explicate linguistic meaning in terms of the relation of a sentence to the mind of a speaker who is using it. To some extent, dynamic semantics (Chierchia (1995)), or rather some versions of it such as Kamp's Discourse Representation Theory (DRT, Kamp (1981)), falls into the latter tradition. At the same time, it remains firmly rooted in a broadly Fregean approach to language. Thus "mind" is not primarily used as a psychological category here, and the focus has been rather on the model-theoretic study of inferential relations between objective propositional contents which are attributed to (and which help to individuate) epistemic states. But the idea of content is given a dynamic twist in that the semantic value of a sentence uttered in conversation is defined to be an update function that expands the body of shared presuppositions of the participants

into a new such body. Thus we have a function,

$$f : C \to C$$

where C is a context, a body of information or a set of presuppositions or propositional attitudes against which a new utterance is interpreted. Semantically speaking, C is a set of possible worlds, but this set is "presented" in terms of a syntactic *discourse representation structure* (DRS). It is something like a formula of predicate logic which is made sensitive to the way a hearer cognitively represent a certain amount of truth conditional information. DRSs are at the same time used as representations of the contents of the beliefs and other attitudes that a hearer forms in response to a discourse.

In this model content and context are unified. A formal semantic representation of a discourse can be looked at as a cont*ent* (a set of truth conditions) as well as a con*text* (for the interpretation of further input). What a sentence is said to "do" is to update the semantic representation generated by the previous sentences, giving a new content as well as context as result. Naturally, this new perspective on meaning involves an increased involvement with pragmatics within semantics. A sentence meaning now itself consists in a dynamic potential for changing a context. It is not merely a semantic content, kept free of the pragmatic elements governing contextual changes.

7.3.2 Context as Common Ground

An empirical hypothesis under which the modelling of multi-speaker discourses in dynamic semantics proceeds, is that

> "speaker and hearer engage in constructing a common ground, maybe from an already existing common ground constructed in earlier conversations" (Zeevat and Scha (1992) p. 25).

Over and above this anthropological hypothesis the meaning of an uttered sentence is typically regarded as

> "an instruction to carry out a series of actions on a given database, with the end effect of incorporating the information given in the sentence into the database" (22).

Notwithstanding certain uses such a modelling decision may have, the question is what it tells us about human communication. Far from being a condition for the possibility of communication, the first above hypothesis restricts the resulting model to the possibly rare type of *common inquiry*, whose participants are unanimous in their goals and values. But agents do not exclusively or mainly strive for the truth, and for getting their findings as effectively as possible across to his fellows.

Even if they engage in common inquiry, agents do not let themselves be "instructed" by whatever is the content that speakers may program into a sentence. What does the meaning of a sentence used in an utterance have to do with the changes I perform on my belief state? It goes without saying that an utterance is a natural object that may have causal effects on interlocutors. But agents in discourses are not generally speaking databases who could be brought to change by laws of cause and effect. How I change my mind is not a matter of my understanding of language, or my linguistic competence (alone). Human agents - and only those are of interest in the present chapter - don't change their mind by being coerced to do so. Understanding is a creative and deliberate process, not a causal one. [6]

It is also argued that for asserting, requesting, promising etc. to take place, interlocutors must attempt to recognize the speaker's intentions, and that this can only be if they cooperate. Speakers, as Grice put it, must

> "make their contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which (they) are engaged" (Grice (1975), 45).

Maybe humans tend to be cooperative as a matter of anthropological fact, but it is unclear whether linguistic theory should engage anthropology in this way. Also, there may not be such an "accepted purpose or direction of the talk exchange", and although an agent may assume or come to the conclusion that there is, he may at any time become uncertain about whether unanimity really prevails. And even if he *does* act cooperatively, we may ask why. That he accepts the Gricean maxims is no answer to this question. We might try to construct an explanation that makes the interlocutor's action come out as an optimal decision in its own best interest, but this would be an explanation within the theory of rational decision. Whether the latter lends itself to explanatory

---

[6]Which also means: it is not a *mechanical* one, as in DRT, where discourse is looked at as a syntactic structure S in which sentences and sets of sentences (discourse segments) are distinguished. If it comes to the interpretation of the i'th sentence S_i, a mechanical device called the *construction algorithm* operates on the syntactic analysis of the sentence and adds its phrasal constituents to the syntactic representation $K_i - 1$ of the discourse up to the sentence $S_i - 1$. The result is a syntactic representation K_i. Philosophically speaking, Fodor's computational theory of thought is operative here, which conceives thought processes as causal-computational relations between mental representations. Note that principled limitations to the Fodorian program argued for by Fodor himself in Fodor (Fodor (2000)) apply, if correct, in full force to the DRT program.

rather than normative purposes is an open question. [7]

A hearer may not bark at some statement that a speaker puts forward, and he may have a good reasons not to do so. He may enter into a political agreement (a professor for example, who listens to the philosophical views of a freshman). Or it may not be useful to weigh the views of others, and to start an inquiry as to whether they are really correct. But it will be some such reasons, which may or may not be there, which explain the fact, if it is one, that a conversation steers towards a common ground, either a political agreement or an actual shared agreement at the end of a common inquiry. In the same vein, we may observe that a *question* is something that an addressee tends to *answer*, although some agents do not. Or that agents take turns in conversation according to some order. But if we want to understand what is the driving force *behind* such phenomena structuring a discourse, explanations come to an end, and normative considerations prevail.

Generally speaking, we make sense of acts of uses of our linguistic competence in the very same way in which we make sense of other kinds of actions, such as the action of playing the piano, which involves use of a musical competence. Musical competence might be scientifically studied, or naturalistically. But a decision to do what we call "play the piano" will likely not, belonging as it does to the realm of human action. The theory of rational action may make this intelligible in some sense, but not necessarily in an explanatory one, as noted. If assumptions concerning common grounds, joint inquiries, cooperativity and so on are to enter linguistic explanation, large questions concerning the explanatory scope of a linguistic theory of this sort arise.

7.3.3   Changing the Context

Robert Stalnaker, one of the forefathers of dynamic semantics, argues that the "contents" of speech acts "should be understood in terms of the way they are intended to affect the situation in which they are performed" (Stalnaker (1998), 4-5). Now it may be that where communicative action is strategic, as in argumentative discourse, my choice of a sentence to make an utterance is a result of considering its potential effects. That however leaves open what this *choice* and its potential contextual effect has to do with the *meaning* that the sentence has in the language in which it is formulated. How speakers change their assumptions in response to an utterance is a process shot through with pragmatic and non-linguistic inferences exploiting background beliefs, social conventions, visual experience of the discourse context, and other

---

[7]Levi Levi (1997) argues forcefully against the explanatory use of the theory of rational decision.

things. While my *choice* of a sentence depends on features of the context, a *sentence* means what it does, in a non-contextual manner, by virtue of the rules of the grammar determining the language. A sentence like

(6) John said to Bill that he wants to wash him

is multiply ambiguous, and this is an essential fact about its meaning. That has to do with the Binding Theory, not with how speakers intend to change contexts.

In short, to the same extent that Stalnaker's "contents" depend on how speakers change contexts through their utterances, they are of dubious value for an analysis of the notion of linguistic meaning. Stalnaker may be right that "[i]t is simply an obvious fact that an assertion changes the context" (Stalnaker (1998), 12), but there is no way to compute, from the sentence used in the assertion, the solution to the question of how the context will change. So what explains the causal effects ensuing from using language?

7.3.4   Discourse as Inquiry

A context in dynamic semantics usually is an *epistemic* context, modeled by a set of assumptions or beliefs, which in turn are analysed as a set of possible worlds in which they are true. If meaning is a context change, it would thus be a matter of the alteration of our propositional attitudes (beliefs, presuppositions). These alterations (context changes) are acts specified under intentional description. That in itself doesn't make the enterprise a normative one. Nonetheless, normative aspects surfaced above for other reasons, and in general, it is hard to see how determing what beliefs a person holds in a context could be an entirely empirical matter. In figuring out beliefs we make assumptions concerning which beliefs it would be *rational* for the person to have in the given context, and that question will guide us to figure out which ones she actually has. Similar considerations apply to the study of epistemic context change, which seems inseparable from the question of how they should change. Hence normativity enters, and meaning, if explicated in terms of the notion of context change, will be to some extent normative, too.

Normativity may quite clearly enter even in the way that considerations about what someone would rationally believe in a context may overrule empirical data about what he says. Somebody who sincerely asserts "I am going downtown" may be interpreted to mean "I am going uptown" if it simply makes no sense at all to interpret him to think or believe that he goes downtown. But note that this shows nothing about

the *meanings* of these sentences, which is just as it should be: the meaning is what it is, by the rules of language, and it doesn't change in the light of considerations of rationality. But context changes do depend on such factors: it doesn't have to do with meaning (alone).

To give an expression to the normativity of epistemic state changes in discourse, we might say that characterizing agents in discourses in terms of their propositional attitudes, as dynamic semantics does, is really to describe them as undertaking certain attitudinal *commitments*. A basis for this consideration might be that forming an attitude is what makes an agent committed to make judgements that correspond to this type of attitude Levi (1997). For example, if an agent comes to the conclusion that a proposition is fully (i.e., non-conditionally, non-partially) true, he undertakes a rational commitment to a disposition to assent to it as well as its logical consequences when prompted to do so.

The shift to talking about beliefs as attitudinal commitments helps to bring out just why discourse updating and context change must be a deliberational matter. Sometimes it may well be the case that you change your commitments routinely in the light of new information. But often you will have to weigh risks against potential gains in changing them: you may be deliberating on whether to trust a person whom who previously distrusted, and while trust may be the option that you would like most and that has the highest value, it might be the most insecure.

Judgements corresponding to the various propositional attitudes allow us to make distinctions between propositions on any chosen level of finde-grainedness. We might distinguish

- propositions that are judged fully true, forming a background for reasoning and inquiry that is not now questioned;
- propositions that are, relative to this background, judged uncertain or probable, where the probability is subjective, and
- propositions that are judged valuable, in accordance with one or the other system of value commitments.

The first propositions I will call (full) *beliefs*, the second *potential beliefs* or *states*, and the third *outcomes*. To indicate how a normative model of discourse updating might get some structure, we might identify the commitment (to the truth of) a proposition with the set of the commitments generated by it, i.e. its (deductive) consequences. This gives us a notion of the 'content' of an attitude, but, of course, a strictly normative, not a semantic one.[8]

---

[8]Levi Levi (1991) adopts such a notion.

154 / Wolfram Hinzen

On this picture, it is not reasonable any more to inquire into the "nature" of propositions. The term proposition as just conceived is not a natural object but a technical notion that has exactly the meaning we give it. What meaning we give it will depend on our purposes: what propositions we wish to distinguish, and how we distinguish them. Independently of given interests and concerns that define what the propositions are that are the objects of a deliberation at a moment of inquiry, and thus when two of them are identical, the question of propositional identity is not meaningful. It cannot be settled by appeal to some absolute authority whether "Hesperus is Hesperus" and "Hesperus is Phosphorus" express the "same proposition". It will depend on the contexts in which we introduce and use these notions that the question will be decided in one or the other way, relative to given purposes and concerns. Language doesn't tell us whether to count them as the same or not, or whether to identify the messages conveyed by "Paul drove Paula home" and "Paula was driven home by Paul".

A major reason why contextuality is massive within rational inquiry is that there may not be general rules for partitioning the space of propositions, or for setting up an ultimate partition of best options that one chooses to consider in the course of one's current decision problem. In a similar way, there seem to be no general rules for determining questions in discourse and for abducing a set of potential answers for them. That is to say: question answering in discourse is not a computational process, but a rational one. There seems to be no clear sense in which the statement that a question is a set of potential answers could be a statement within a descriptive linguistic theory. Not even the construct of a shared language that I argued against above can make the deliberative and normative aspects in constructing and answering a question go away: language doesn't have the effect to make you construct, by itself, a set of potential answers relevant to the speech situation. Consider the interrogative

(7) Do you accept Credit Cards?

spoken by me when calling a restaurant. It does not come with a partition of potential answers, nor does it determine, by itself, what the question is in the first place. It may not in fact be a question, but, for example, an overt or hidden threat. If it is a question, it may be inferred that I want to know whether one of the accepted Credit Cards is the one I own; whether I can pay with one of my Credit Cards, which are most probably among those accepted in a restaurant; whether I can patronize this restaurant; etc. Having inferred the question, a Hearer abducts a set of strongest consistent, exhaustive, and pairwise incon-

sistent potential answers. The final outcome of this abductive phase of deliberation contains those potential answers that are the options between which one has decided to restrict one's choice. It is exactly as fine-grained and as coarse-grained as one finds it useful. For all it seems, no general rules for forming a partition can be given. In a similar way, it seems that no general rules can be given for what demands of information an agent should have, or with what degree of caution he should reject hypotheses.

In sum, what I have tried to bring out is that what dynamic semantics really seems inspired by is a vision of *inquiry*, not meaning, and that eventually that vision will have to show its true colours as a *normative* one. We might indeed define a technical notion of "meaning" so as to consist in the set-difference of the Hearer's partition before and after the utterance of a sentence has taken place. In this way we would read off the meaning of a sentence from the inductive inference (an elimination of options) that an agent performs. That this is a non-linguistic notion of meaning seems rather clear, however, and it might be more useful to call it a notion of *significance*. In any case, its *massive contextuality* - the fact that it will depend on contextually given human interests and values in an almost arbitrary manner - makes it an unlikely candidate for a naturalistic inquiry into human nature and language.

7.4   Stalnaker's Critique of Dynamic Semantics

Dynamic semantics is often said to have one of its origins in Stalnaker's work on assertion. Nonetheless Stalnaker (Stalnaker (1998)) criticizes it precisely for taking the pragmatics of contextual change to define *semantic* content. The attack recommends separating a purely semantic core notion of meaning from pragmatic aspects of meaning having to do with the use of expressions that have that semantic content. Linguistic phenomena that supposedly motivate the introduction of dynamic semantic values (functions from contexts to contexts), Stalnaker argues, can be accounted for just as well with classical models of contexts and propositions (sets of possible worlds). It is the *point* of a meaningful speech act to change a context; but this is not what defines the *meaning* of a sentence.

All this is intendedly Gricean in spirit: Grice also reconciled the traditional non-contextual semantic analyses of logical expressions by adding a non-semantic account of conversational maxims to them. These would leave the assigned semantic contents untouched but lead us to the specific interpretations that a context might require. But in

the case of Stalnaker we may ask exactly what we may ask in the case of Grice: whether the notion of semantic content that we protect by adding pragmatics to the theory of language is an artificial phenomenon rather than a natural kind.[9] Both the notion of proposition and the notion of meaning just used are fictions if the above discussion is right: they are artificial abstractions from the massively contextual interplay of a myriad of faculties as they enter public discourse. We will now try to substantiate this prior conclusion above in the light of Stalnaker's account.

It is essential to Stalnaker's argument that the proposition, in his sense, that is asserted (or "proposed for acceptance", as Stalnaker says) is fixed before we look at the contextual dynamics that the assertion of the sentence expressing that proposition affects. Specifically, if a context changes through the assertion of a sentence, it does so in two ways: first the set of mutually shared assumptions adjusts to the fact that a *particular sentence with a particular content* has been asserted (first context change); secondly, the proposition proposed for acceptance is either accepted or rejected (second context change) (Stalnaker (1998), 8). Semantics is dealt with when the first change has happened. Pragmatic principles control the second. Both changes are formally modelled as modifications of sets of possible worlds modelling the information that is compatible with the shared assumptions of a group of speakers at a particular moment of their discourse.

Note that Stalnaker's second step of contextual change involves adopting or not adopting a belief that will in general not be a logical consequence of what is believed already. In this sense the inference that is drawn to its truth is ampliative or inductive. It will necessarily carry risk in the sense alluded to above. An inference that proceeds from a body of beliefs that are held true to a new such body involves a risk of error from the viewpoint of what is currently held true. While it is trivial to model such an induction as a change in the set of possible worlds capturing what is taken for granted at a stage in discourse, what is decisive in a theory is the reasons why such changes happen. One cannot built the rational structure and the human values that enter into the reasons for such changes into a number of "conversational maxims" in

---

[9] As is sometimes said: Semantics was saved by inventing pragmatics. But what if there is no semantics? It is not clear whether there is not, but it seems necessary that with respect to natural languages the existence of a level of semantics and "propositional contents" is an empirical matter entirely. See e.g. Hornstein Hornstein (1983) for general arguments that there is no such level, Jennings Jennings (1995) for empirical refutations of Gricean stipulations of logical constants in natural language, and Dudman Dudman (1991) for the specific case of English *if*.

Grice's manner, or, as Stalnaker's expresses this, a "number of truisms about conversation as a rational activity" (Stalnaker (1998), p. 18).

Let us look at the fabrication of Stalnaker-propositions in some detail. Recall Stalnaker's first type of context change, caused by the "manifestly observable event" (Stalnaker (1998), 8) that a "statement was made". But together with it and "standing assumptions" that are part of the prior context we can "infer, (...) not only that the speaker uttered certain sounds, but also that she uttered an English sentence, and that she is saying something to us" (8-9). Further assumptions (cooperativity, competence)- Stalnaker goes on to assume - make inferable from observational data, in fact, that we have a token *of an English sentence type whose meaning and presuppositions in the event of its utterance we know.* For example, when Phoebe asserts

(8) I can't come to the meeting - I have to pick up my cat at the veterinarian,

the first context change that takes place according to Stalnaker is that we now presuppose that Phoebe owns a cat, and that this is now a shared presupposition.

But why do we do such a thing? Stalnaker answers: because after we witness the utterance of (8), we presuppose, not only that it took place, but also that Phoebe "is making whatever presuppositions are required to make her utterance intelligible and appropriate" (9). This in turn is because we have the prior belief that Phoebe is a "cooperate and competent participant in our conversational enterprise" (9). From these assumptions, together with the further one that the expression *my cat* has as part of its "appropriateness conditions" that its speaker owns a cat, we presuppose that Phoebe means to presuppose that she owns a cat, and presumes this information to be shared. Finally, "since it is presupposed that presuppositions are shared information, he will accommodate by presupposing it himself" (10).

It should be clear that all of this takes a lot of good will. In fact it seems that if we ask again why we do any of these things, what we are pointed to is the fact that there are certain "standing assumptions", both about what Phoebe is like, and what it is for certain expressions to be used "appropriately" (note how the shared language assumptions is presupposed here). From this it would seem that whatever is needed to get us to draw all those inferences is coded into our premises. Be that as it may, it remains that all of the inferences and conclusions mentioned that we draw from an utterance event are *inductive* ones in the sense above: but that by itself implies that we *need* not draw them. It is not enough to point out that we will "typically" draw them, or that

there is a "disposition" to do so. Drawing an inductive inference is to go beyond present commitments, and this is and remains a deliberative matter.

To sum up this discussion, it seems that at the point where we have got a Stalnaker-proposition (an absolute, non-contextual item of information, a set of possible worlds) and a presupposition associated with it, massive inferencing has already taken place. When it comes to Stalnaker's second kind of context change - where a proposed proposition is accepted or rejected - what we have is a product of this second kind of context change (a belief change, an either routine or deliberate expansion, in the light of an utterance event) already. The determination of the proposition proposed for acceptance in a speech act itself requires making reference to pragmatic and inductive principles for changing one's context or doxastic dispositions. This is evident even given the assumptions Stalnaker makes in his example. It becomes abundantly clear if the shared language assumption fails, or the inductively inferred cooperativity and competence assumptions are given up. Semantic content in Stalnaker's sense is itself largely a matter of pragmatics and inductive inference, and not given independently by "language" itself.

Stalnaker himself in fact remarks that we may happen not to be "willing" to infer that Phoebe has a cat. What is surprising is that a discourse should have to be rated "defective" (Stalnaker (1998), 10) if this is so (it would appear that such clearly normative considerations have no place in semantic theory). Due to the defectiveness, Stalnaker argues, some "backtracking and repair" is required. We may conclude from this that even for Stalnaker, in the *general* case how a context changes in the first sense is a matter of pragmatics and reasoning, hence is itself a context change in his second sense. It seems then that Stalnaker has no point in claiming that dynamic semantics mixes pragmatic matters into its notion of a semantic value. Stalnaker himself doesn't isolate a clearly semantic notion of linguistic content that would be logically prior to the context changes that the dynamic semanticist talks about.

Let us finally apply these conclusions to a concrete and famous example of Partee's by virtue of which Stalnaker aims to show that dynamic semantic values aren't needed. Imagine the following well-known pair of examples, said in the same initial context:

(9a) Exactly one of the ten balls is not in the bag.
(9b) Exactly nine of the ten balls are in the bag.

One argument for dynamic semantic values went as follows: If the contents of (9a) and (9b) are modelled as sets of possible worlds, their

contents will come out as exactly the same. This, dynamic semantics has claimed, is a problem since if we continue a discourse started with either sentence with the new sentence

(10) It is under the sofa

this sentence will be interpreted differently in the two cases. The possible world analysis doesn't predict this.

The supposed problem for the possible worlds analysis vanishes, Stalnaker argues, if his first kind of context changes is taken seriously. For since different words were used in the utterances of the respective sentences (9a) and (9b) to get across what is in some technical sense the "same content", the context will change in different ways. An utterance of (9a) makes an antecedent for *it* in (10) uniquely salient, and (9b) fails to do so. This is a fact that "can be presumed to be available" to the speakers, hence, if sets of possible worlds model the context of interpretation prior to the utterance of (10), these sets will be different in the case of (9a) and (9b).

While that may be right, it leaves open, as Stalnaker acknowledges, *why* they are different. But this is what we would like to explain. The set difference is no more than an expression of what we wish to understand. And there is a problem of understanding, for, as noted, it simply won't do to say that a sentence is a context changer. Sentences, as noted, do not do such things, and even if *it* in (10) picks up the one ball made salient through an utterance of (9a), it is not facts about (9a) qua English sentence which cause this. Neither does it help to stipulate that "the point of a speech act (...) is to change the context" (1998, 8). Stated as such this cannot be right, as it can only be the point of an assertion to change the context *in a certain way*. But if thus understood, we derive what is to be explained by introducing a new entity, a communicative intention, which fixes the point of an assertion. But the phenomenon of the specific context changes we have to explain *consists in* differences relating to making a certain communicative act, and we cannot appeal to the latter to explain the former.

Maybe we could call the process that leads to assign the missing ball to *it* a default inference. If so, a default inference is one that needs to be *drawn* by a hearer, and it is again not clear whether this is a matter of linguistic form alone. The assignment of a referent to *it* in the case of an utterance of (9b) - namely, an assignment of either the bag or the missing ball (which might as well have been made salient by mentioning the ones that are *there*) - is in either case clearly not an inference based on linguistic form. Of course there is nothing wrong in inferences not based on linguistic form, but an explanation won't

benefit if facts about what inferences we may or may not draw from linguistic facts are packed into our notion of linguistic meaning.

Even if the assignment of the missing ball to *it* in the case of an utterance of (9a) were a linguistic matter, a postulation of richer semantic representations than is offered by Stalnaker's austere apparatus of possible worlds might be very useful. Thus Kamp's DRT suggests formal semantic representations for (9a) and (9b) which differ in that one contains a discourse referent which subsequent pronominal discourse may pick up, while the other does not. Here the existence of this difference in formal structure is a way of *explaining* the set difference that Stalnaker mere takes as given. Furthermore, DRT offers a specific construction algorithm operating on the syntactic structures of the sentences in (9) which yields the difference in formal semantic representation. In other words, the formal difference in representation and content is traced to linguistic mechanisms, not invited inferencing. This is good because it doesn't let semantic contents be made up of non-linguistic facts such as invited pragmatic inferences. It is bad if I am right that the inferences leading to an assignment of a referent to *it* both in the case of an utterance of (9a) and (9b) are of a non-linguistic sort, in which case a non-linguistic inductive inference would be wrongly modelled as a linguistic mechanism.

I am not insisting on a point that boils down to the question how we define language. The question how rich a notion of language we can afford if we are after an explanatory theory of language. Suppose it was right that

> "speech acts are context-dependent: their contents (and so the way they are intended to affect the situation) depend not only on the syntactic and semantic properties of the types of the expressions used, but also on facts about the situation in which the expressions are used" (Stalnaker (1998), 5).

Then it would be true that computations that the mind performs when constructing a semantic representation of an input sentence cannot be *local* in a sense suggested to be necessary for a formal account of human knowledge of language in the beginning of this paper. If the notion of content employed in the quote was adopted, principles for computing a semantic representation couldn't be defined locally over features represented in the human head, let alone an encapsulated language module. If the non-linguistic environment in which expressions are used relevantly enters the computation, it will be dubious whether we can still be talking about a theory of semantic representation as

part of a computational theory of cognition.[10]

I finally and briefly mention another problem in the Stalnakerian picture, relating to questions of methodology. The strategy is "to get clear about what language is for - what it is supposed to do - before explaining how it does it" (Stalnaker (1998), 4). But language is a natural phenomenon, and it seems in general dubious to ask for a natural phenomenon "what it is for". Language is used to change contexts for sure, but the question is whether this teleological talk is more than a matter of historical narrative. Communicative purposes may be invoked in informal description of the biological phenomenon, but abandoned as we progress to explanatory theories, in which we discard notions of function, intention, and purpose. The concrete "mechanisms and devices that particular languages may provide for doing what is done in a discourse" appear to be what we ultimately aim for, and I know of no reason to believe that these mechanims will have to vindicate our common sense intentional description of "doing what is done in a discourse", or even be understood as means to achieve such doings. There remains a question, then, about the status of the "descriptive apparatus" that Stalnaker proposes to develop to study the interaction of "content and context". If we study a natural phenomenon, maybe we should aim not for some such "apparatus", but an explanatory framework within which we can try to find real categories of human cognition.

7.5 Contexts in LFs

7.5.1 Context Variables

The notion of meaning employed in section 2 above was a non-contextual one, leaving matters of interpretation relating to the utterances of expressions aside. I granted that in the interpretation of utterances contextual factors of any imaginable sort are likely to enter the process of interpretation. I will now ask whether there are linguistic ones among those factors.

Assuming it is not words which refer to things but people which do that on occasions of using them, we can start by noting a familiar contextuality with respect to indexicals like *I* and demonstratives like *this*. One says an expression like *I* is "context-dependent", meaning the referent shifts on who uses the expression. In the case of *I*, the very act of using the word is sufficient to fix the referent. But that should be

---

[10]See Fodor Fodor (2000) for a systematic elaboration on the locality properties needed for cognitive processes to figure in a computational theory. Note again that if dynamic theories of discourse took Fodor's point seriously, they would have to exclude abductive processes - which Fodor argues are non-local - from their account of semantic representation.

said to be a feature of the word *I* relating to *non-contextual* features of its meaning, for what is contextual is the referent determined by features of its meaning in an act of using it.

Similarly, what is non-contextual about the word *this* is that it must be accompanied by an act of reference - a speaker must use it to refer to something or other - even though only the context of use will tell what object is being referred to. A pronoun like *she* would appear to be intermediate between these extremes, constraining on linguistic grounds the referent to be a person that is female, but leaving the choice of actual reference within those confines to the speaker.

In line with these observations concerning the linguistic and the pragmatic (or use-theoretic) aspects of an expression which enter interpretation, Higginbotham Higginbotham (1988) suggests stating truth conditions for context-dependent utterances like

(10) She is lazy

in the following manner.

(11) If x is referred to by *she* in the course of an utterance of (11), and x is female, then that utterance is true just in case lazy(x).

As Higginbotham notes (p. 34), this makes the truth apply to utterances in an 'absolute' fashion: there is no notion of truth-in-a-context in (11). Contextual features that influence the act of interpretation are mentioned in the conditional's antecedent, and the right hand side of the biconditional in the conditional's consequent registers their effects on the interpretation.

The semantics in the style of (11) needs further refinement for cases in which a speaker doesn't use a noun phrase to refer to a thing but merely restricts the range of things that the phrase may refer to. Examples are (12), on the reading where *he* is bound by *everyone*, and the famous (13):

(12) Everyone likes the book he read
(13) The murderer (whoever he is) is insane.

Even though it is required by the linguistic form of (12) alone that everyone read a single object and that this object was a book, otherwise the referent is indeterminate. Inspection of the context of an utterance of (12) will provide the referent(s). Similarly in (13), where no particular person is referred to. Rather, the noun *murderer* is used to restrict the quantification to persons who murdered someone, and the speaker's role is, in Higginbotham's phrase, to *confine the range of the noun to things such that* certain contextually given conditions are fulfilled. Symbolizing these conditions by a free, second-order variable X serving to encode

contextual restrictions on the range of quantification of expressions with generality impact, the normal form of stating truth conditions for something like (13) becomes:

(14) In any utterance of (13) in which the speaker confines the range of *murderer* to things x such that Xx, that utterance is true just in case: insane$(($the $x)($murderer(x) and $Xx))$

Context variables will have to be employed in full generality, as it is standardly the case that ranges of quantification of general expression are usually not their full ranges but appropriated to a context. In capturing contextual effects on interpretation in this manner we are not delving into a study of the multifarious facets of communicative contexts, but specify what is required, on the linguistic side of the interface with the pragmatic systems, to determine the content of an utterance.

A significant virtue of context variables is manifest in the semantics of names. As Uriagereka Uriagereka (1998) notes in the context of other examples, there is nothing wrong, say, in accepting that while Pizza Quattro Stagioni tastes great in my local pizza place, it doesn't taste great today (considering the instance here in front of me). How is that consistent (it obviously is)? Apparently, by noting that human languages allow us to distinguish between individuals (such as pizzas) *at* certain events in which they participate (such as the current eating event), on the one hand, and self-standing or non-contextual individuals, on the other. Obviously, we are ill-advised to follow traditional practice and formalize names as simple constants denoting or labeling non-contextual objects in the world. It may be essential to the objects as referred to by us that they occur in a certain context in which they play a certain role. The pizza Quattro Stagioni in today's dining event is a different Pizza Quattro Stagioni from the one I know and love. It seems like a crucial fact of human cognition that it is in contexts that speakers present notions in specific respects.[11] In the next section I present observations of Raposo and Uriagereka Raposo and Uriagereka (1995) that, furthermore, it is in contexts that speakers introduce *other*

---

[11]From this point of view there is a clear disagreement with Stanley's claim (Stanley (2000), 400) that overt expressions that are neither indexicals nor demonstratives are "not context-dependent". If what I said is right, there is a sense in which we would rather have to say that literally every expression, no matter its category, is "context-dependent". There is no fixation of reference without, for example, taking speaker's non-linguistic intentions and presuppositions into account, even in simplest cases. On the other hand, given the aboce context variables one sees the affinity to Stanley's approach in his formulation that "any contextual effect on truth-conditions that is not traceable to an indexical, pronoun, or demonstrative (...) must be traceable to a structural position occupied by a variable" (Stanley (2000), 401).

contexts, hence that context variables take scope over one other, much as quantifiers do.

7.5.2 Relative Scope of Context Variables

The way context variables are *ordered* with respect to each other appears to account for differences in predication that can be observed in examples like (15):

(15a) The city has poor neighboorhoods
(15b) The city is (one) of poor neighboorhoods.

There is an obvious semantic difference here, relating to the fact that in (16a), we consider the city in the context of a particular issue, namely poor neighborhoods. With respect to that, the city has those neighboorhoods, but nothing is said as to whether it has rich ones as well. That is not so in (16b), where the poor neighboorhoods are what we might call a "standing characteristic" of the city, which holds of it in a decontextualized fashion. In general terms, while in (15a), the context of the city is grounded on that of the predicate, it is the other way round in (15b). The ambiguity can be observed generally, as in

(16) Marlowe loved women

which can be interpreted as a statement about a standing characteristic of Marlowe, a factual feature of his personality, but equally as a report on an erotic event or a sequence of such events in Marlowe's life. In the latter case, we might say that it's the loving of women that we talk about, and which figures as the sentence's topic.

Such observations suggest that assuming every quantificational expression comes with a context, and assuming further that each predicate comes with an extra event argument Higginbotham (1983), each simple "S is P" predication leaves it open which context is set into which. Either the predicate is topicalized (moved to a site marked for topicality at SEM), in which case we talk of some event in the context of which we consider some particular thing as its subject. Or the subject is topicalized, in which case we introduce the predicate in the context of that fully individualized object, of which it will now hold as a standing characteristic. A striking illustration of Raposo and Uriagereka Raposo and Uriagereka (1995), who suggest this idea, is (17) and (18):

(17) I consider the seathe frog green
(18) I saw the seathe frog green.

There is a canonicity effect with respect to the greenness of the sea and the frog, respectively, in (17): the sea green will very possibly differ from the frog green (the former being, say, more blueish). That

typicality of the colour is not invoked in the case of (18), where as a result of the light and visual circumstances we may well have seen both the sea and the frog in exactly the same green.[12] We can explain these observations in the light of the prior considerations. If the green is predicated in the context set up by the subjects *sea* and *frog* considered as topics, it is the green which is canonical to these items. If, on the other hand, what we talk about is an event of seeing greeness, and the seafrog are as it were reduced to participants in this event, the green that we see may be whatever it is: the sea and the frog will have *it*, and it will not in any essential way be the sea's green or the frog's green.

In consequence, there is evidence that it is not merely that context variables are instantiated in context on the basis of whatever information and cues are contextually given, they are also instantiated sequentially one *after* the other in the process of constructing semantic representations.

7.6 Conclusions

The emerging picture appears to be this. I have started arguing that a theoretically promising notion of language must be wary against the contextuality that is relevant and pervasive to the study of human communication and purposive language use. While it is true that we use language to communicative purposes and the transmission of thoughts, this need not explain us why languages generate sructures and meanings in the way they do. Linguistics on this picture is a science of expressions, not a study of actions specified under intentional description, hence completely remote from the normative issues that unavoidably arise in the latter.

This specifies the role of the non-linguistic context given by the uses to which expressions are put. Semantics, if conceived as the study of rules by which linguistic sounds relate to meanings, is non-contextual. On occasions of use, matters of linguistic form are contextualized, but the massive contextuality found in inductive inferencing does not seem to lead to a foundational clear and fruitful notion of content. On the other hand, none of this hinders variables encoding context to enter linguistic description and to influence our conception of LF, and it even may give rise to reconsidering certain traditional conceptions in the theory of object and event reference.

---

[12]Data about auxiliary selection in the Spanish counterparts to these examples make the difference of predication overtly manifest in the syntactic form.

166 / Wolfram Hinzen

References

Chierchia, G. 1995. *Dynamics of Meaning*. Chicago: University of Chicago Press.

Chomsky, N. 1995. *The Minimalist Program*. Cam., Mass.: The MIT Press.

Chomsky, N. 2000. *New Horizons in the Study of Language and Mind*. Cambridge University Press.

Clark, H. 1992. *Arenas of Language Use*. Chicago: University of Chicago Press.

DiSciullo, A.M. 1999. Formal Context and Morphological Analysis. In P. Bouquet, L. Serafini, P. Brezillon, M. Benerecetti, and F. Castellani, eds., *Modeling and Using Context*, vol. 1688 of *Lecture Notes in Artificial Intelligence*, pages 105–118. Springer Verlag.

Dudman, V. 1991. Interpretations of 'if'-sentences. In F. Jackson, ed., *Conditionals*, pages 202–232. Oxford University Press.

Fodor, J. 1998. *Concepts*. Oxford: Clarendon.

Fodor, J. 2000. *The mind doesn't work that way*. Cam., Mass.: The MIT Press.

Grice, P. 1975. Logic and Conversation. In M. J. Cole P., ed., *Syntax and Semantics*, vol. 3: Speech Acts. New York: Academic Press.

Higginbotham, J. 1983. The Logic of Perceptual Reports. In *Occasional Papers from the Center for Cognitive Science*. MIT.

Higginbotham, J. 1988. Contexts, models, and meanings. In R. Kempson, ed., *Mental Representations*, pages 29–48. Cambridge UP.

Hinzen, W. 2001. The pragmatics of inferential content. *Synthese* 128(1–2):157–181.

Hornstein, N. 1983. *Logic as Grammar*. Cam., Mass.: MIT Press.

Huang, J. 1995. Logical Form. In G. Webelhuth, ed., *Government and Binding Theory and the Minimalist Program*, pages 125–176. Oxford: Blackwell.

Jennings, R. 1995. *The Genealogy of Disjunction*. Oxford: Oxford University Press.

Kamp, H. 1981. A Theory of Truth and Semantic Representation. In J. Groenendijk, T. Janssen, and M. Stokhof, eds., *Formal Methods in the Study of Language*, pages 277–322. Amsterdam: Mathematical Centre.

Kamp, H. and U. Reyle. 1993. *From Discourse to Logic*. Kluwer.

Levi, I. 1991. *The Fixation of Belief and its Undoing*. Cambridge: Cambridge University Press.

Levi, I. 1997. *The Covenant of Reason*. Cambridge: Cambridge University Press.

Nunberg, G. 1995. Transfers of Meaning. *Journal of Semantics* 1, Winter.

Raposo, E. and J. Uriagereka. 1995. Two Types of Small Clauses (Toward a Syntax of Theme/Rheme Relations). In A. Cardinaletti and M. T. Guasti, eds., *Syntax and Semantics*, pages 179–206. Academic Press.

Stalnaker, R. 1978. Assertion. *Syntax and Semantics* 9: Pragmatics.

Stalnaker, R. 1998. On the Representation of Context. *Journal of Logic, Language and Information* 7:3–19.

Stanley, J. 2000. Context and Logical Form. *Linguistics and Philosophy* 23:391–434.

Uriagereka, J. 1998. A Note on Rigidity. *Linguistics Today* 22: Possessors, Predicates and Movement in the Determiner Phrase:361–382.

Zeevat, H. and R. Scha. 1992. Integrating Pragmatics into Update Semantics. In S. O. Ortony A., Slack J., ed., *Communication from an Artificial Intelligence Perspective*, pages 17–34. Springer.

March 22, 2006

8

Truth-conditional pragmatics: an overview

Franois Recanati

8.1 Two kinds of interpretation

Most philosophers of language, and many linguists, view the division of labour between semantics and pragmatics as follows. Semantics deals with the literal meaning of words and sentences as determined by the rules of the language, while pragmatics deals with what users of the language mean by their utterances of words or sentences. To determine 'what the speaker means' is to answer questions such as: Was John's utterance intended as a piece of advice or as a threat? By saying that it was late, did Mary mean that I should have left earlier? Notions such as that of illocutionary force (Austin (1975)) and conversational implicature (Grice (1989)) thus turn out to be the central pragmatic notions. In contrast, the central semantic notions turn out to be reference and truth. It is in terms of these notions that one can make explicit what the conventional significance of most words and expressions consists in. So the usual story goes.

As Grice emphasized, speaker's meaning is a matter of intentions: what someone means is what he or she overtly intends (or, as Grice says, 'M-intends') to get across through his or her utterance. Communication succeeds when the M-intentions of the speaker are recognized by the hearer. Part of the evidence used by the hearer in figuring out what the speaker means is provided by the literal meaning of the uttered sentence, to which the hearer has independent access via his knowledge of the language. In ideal cases of linguistic communication, the speaker

means exactly what she says, and no more is required to understand the speech act than a correct understanding of the sentence uttered in performing it. In real life, however, what the speaker means typically goes beyond, or otherwise diverges from, what the uttered sentence literally means. In such cases the hearer must rely on background knowledge to determine what the speaker means - what his or her communicative intentions are - on the basis of what he or she actually says. On this view two distinct and radically different processes are jointly involved in the interpretation of linguistic utterances: semantic interpretation and pragmatic interpretation. They are standardly described as follows:

- Knowing a language is like knowing a theory by means of which one can deductively establish the truth-conditions of arbitrary sentences of that language. Semantic interpretation consists in applying that theory to a particular sentence of the language so as to determine its truth-conditions on the basis of the references of its parts and the way they are syntactically combined.

- Pragmatic interpretation is a totally different process. It is not concerned with language per se, but with human action. When someone acts, whether linguistically or otherwise, there is a reason why he does what he does. To provide an interpretation for the action is to find that reason, that is, to ascribe the agent a particular intention (for example, a communicative intention) in terms of which we can make sense of the action.

Two important characteristics of pragmatic interpretation, as opposed to semantic interpretation, stand out and must be stated from the outset.

First, pragmatic interpretation is possible only if we presuppose that the agent is rational. To interpret an action, we have to make hypotheses concerning the agent's beliefs and desires; hypotheses in virtue of which it can be deemed rational for the agent to behave as she does. This feature of pragmatic interpretation I will refer to as its *hermeneutic* character. It strikingly contrasts with the algorithmic, mechanical character of semantic interpretation (as standardly conceived).

Second, and relatedly, pragmatic interpretation is always *defeasible*. The best explanation we can offer for an action given the available evidence may be revised in the light of new evidence. Even if an excellent explanation is available, it can always be overriden if enough new evidence is adduced to account for the subject's behaviour. This, again, contrasts with the non-defeasible, monotonic character of semantic interpretation.

A third contrast worth stating concerns the role of *context* in se-

mantic and pragmatic interpretation. Because of its defeasibility - what Stainton (J.Stainton (2005)) calls its 'all-things-considered' character - there is no limit to the amount of contextual information that can in principle affect pragmatic interpretation. But context comes into play in semantic interpretation only to help determine the reference of those few expressions whose reference is not fixed directly by the rules of the language but is fixed by them only 'relative to context'. The context at issue is a small package of factors involving only very limited aspects of the actual situation of utterance: who speaks, when, where, to whom, and so forth. In contrast, the context relevant to determining what the speaker means is all-inclusive. Any aspect of the total world in which the utterance takes place (not to mention the 'possible worlds' projected by the beliefs, intentions, etc., of the language users) is part of the context which can affect pragmatic interpretation. In a nutshell: *Anything* can affect pragmatic interpretation (as opposed to semantic interpretation, which is 'informationally encapsulated'[1].

From what has just been said, it follows that there are two notions of context: a narrow and a broad one, corresponding to semantic and pragmatic interpretation respectively. As Kent Bach puts it:

> Wide context concerns any contextual information relevant to determining the speaker's intention and to the successful and felicitous performance of the speech act... Narrow context concerns information specifically relevant to determining the semantic values of [indexicals]... Narrow context is semantic, wide context pragmatic.[2]

When the (narrow) context comes into play to determine the semantic values of indexicals, it does so in the algorithmic and non-hermeneutical manner which is characteristic of semantic interpretation as opposed to pragmatic interpretation. The narrow context determines, say, that 'I' refers to John when John says 'I', quite irrespective of John's beliefs and intentions. As Barwise and Perry write, "even if I am fully convinced that I am Napoleon, my use of 'I' designates me, not him. Similarly, I may be fully convinced that it is 1789, but it does not make my use of 'now' about a time in 1789" (Barwise and Perry, 1983, p. 148).

---

[1] Informational encapsulation is one of the characteristic features of cognitive 'modules'. Such modules do not "have access to all of the information that the organism internally represents" (Fodor, 1983, p. 69), but only to a restricted range of data.

[2] From the handout of a talk on "Semantics vs. Pragmatics" delivered in 1996.

8.2   What is wrong with the traditional view

According to the traditional view I have just expounded[3], knowledge of
a language enables interpreters to deductively assign truth-conditions
to arbitrary sentences of that language. To be sure, if the sentence
contains an indexical expression, its truth-conditional content will de-
pend upon the context; mere knowledge of the language will not be
sufficient for truth-conditional interpretation. But the context relevant
to content-determination in such cases is the narrow context. To fix
the truth-conditional content of even indexical utterances, there is no
need to engage in 'all-things-considered' reasoning, no need to take the
speaker's beliefs and intentions into consideration. *That* is the gist of
the traditional picture. Knowledge of the language and, if necessary, of
the (narrow) context suffices for truth-conditional interpretation. Con-
siderations pertaining to the speaker's beliefs and intentions become
relevant only when we are to determine what the speaker means by or
in saying what she says. In other words: *semantic interpretation de-
livers truth-conditions; pragmatic interpretation determines aspects of
utterance meaning over and above truth-conditions.*

This picture has come under sustained attack during the last fif-
teen years, and an alternative picture has been put forward: Truth-
conditional pragmatics (TCP). The gist of the new picture is that
we need pragmatic interpretation to get truth-conditions in the first
place. Pragmatic processes are involved not only to determine what
the speaker means on the basis of what she says, but *also* to determine
what is said, insofar as this is distinct from the meaning of the sentence
type.

Recall that, on the traditional view, the reference of indexicals is
determined automatically on the basis of a linguistic rule, without tak-
ing the speaker's beliefs and intentions into consideration. Now this

---

[3]In an earlier version of this chapter I had called the view in question the 'stan-
dard view'. But Andrea Bonomi raised the following objection: "If I look at the
specific analyses developed in the area of truth-conditional semantics (in its model-
theoretic version) what I see is a systematic reference, in different forms, to the *wide*
notion of context where what is crucial is a common ground of intentions and beliefs.
This is so, for example, in most treatments of quantified NPs, in the semantics of
counterfactuals, in the analysis of the processes of "updating" developed in dynamic
semantics, or in several treatments of aspectual features . . . If this is correct, then
the very notion of a "standard view" (characterized by the idea that 'knowledge of
the language and, if necessary, of the narrow context suffices for truth-conditional
interpretation') might be based upon an extreme, perhaps excessive, simplification,
because what semanticists really do, in their concrete work, cannot be identified
with that view". I agree with Bonomi that the traditional view, though still promi-
nent in the official ideology (especially among philosophers), is hardly consistent
with the actual practice of semanticists.

may be true of some of the expressions which Kaplan classifies as 'pure indexicals' Kaplan (1989), but it is certainly not true of those which he calls 'demonstratives'. The reference of a demonstrative cannot be determined by a rule, like the rule that 'I' refers to the speaker. It is generally assumed that there is such a rule, namely the rule that the demonstrative refers to the object which happens to be demonstrated or which happens to be the most salient, in the context at hand. But the notions of 'demonstration' and 'salience' are pragmatic notions in disguise. Ultimately, a demonstrative refers to *what the speaker who uses it refers to by using it.*

To be sure, one can make that into a semantic rule. One can say that the 'character' of a demonstrative is the rule that it refers to what the speaker intends to refer to. As a result, one will incorporate a sequence of 'speaker's intended referents' into the narrow context, in such a way that the nth demonstrative in the sentence will refer to the nth member of the sequence. Formally that is fine, but philosophically it is clear that one is cheating. We pretend that we can manage with a limited, narrow notion of context of the sort we need for handling pure indexicals, while in fact we can only determine the speaker's intended referent (hence the narrow context relevant to the interpretation of the utterance) by resorting to pragmatic interpretation and relying on the wide context.

We encounter the same problem even with expressions like 'here' and 'now' which Kaplan classifies as *pure* indexicals (rather than demonstratives). Their semantic value is said to be the time or place of the context respectively. But what counts as the time and place of the context? How inclusive must the time or place in question be? It depends on what the speaker means, hence, again, on the wide context. We can maintain that the character of 'here' and 'now' is the rule that the expression refers to 'the' time or 'the' place of the context - a rule which automatically determines a content, given a (narrow) context in which the time and place parameters are given specific values; but then we have to let a pragmatic process take place to fix the values in question, that is, to determine *which* narrow context, among indefinitely many candidates compatible with the facts of the utterance, serves as argument to the character function. On the resulting view the (narrow) context with respect to which an utterance is interpreted is not given, it is not determined automatically by objective facts like where and when the utterance takes place, but it is determined by the speaker's intention and the wide context. Again we reach the conclusion that, formal tricks notwithstanding, pragmatic interpretation has a role to play in determining the content of the utterance.

The alleged automaticity of content-determination and its indepen-

dence from pragmatic considerations is an illusion due to an excessive concern with a sub-class of 'pure indexicals', namely words such as 'I', 'today' etc. But they are only a special case - the end of a spectrum. In most cases the reference of a context-sensitive expression is determined on a pragmatic basis. That is true not only of standard indexical expressions, but also of many constructions involving something like a free variable. For example, a possessive phrase such as 'John's car' arguably means something like *the car that bears relation R to John.* The free variable '*R*' must be contextually assigned a particular value; but that value is not determined by a rule and it is not a function of a particular aspect of the narrow context. What a given occurrence of the phrase 'John's car' means ultimately depends upon what the speaker who utters it means. It therefore depends upon the wide context. That dependence upon the wide context is a characteristic feature of 'semantically underdetermined' expressions, which are pervasive in natural language. Their semantic value varies from occurrence to occurrence, yet it varies not as a function of some objective feature of the narrow context but as a function of *what the speaker means.* Pragmatic interpretation is therefore necessary to determine what is said by a sentence containing such an expression.

8.3   Primary pragmatic processes

The pragmatic processes that are involved in the determination of what is said, and which justify talk of 'truth-conditional pragmatics', I call *primary* pragmatic processes. *Qua* pragmatic processes, they have all the properties characteristic of pragmatic interpretation: in particular, defeasibility and dependence upon the wide context. According to truth-conditional pragmatics, when an utterance is made and a certain truth-conditional interpretation emerges for that utterance, it does so as a result of pragmatic processes that can be affected by any change in the wide context. A given truth-conditional interpretation for an utterance can therefore always be revised in the light of additional background information.

There are two classes of primary pragmatic processes: top-down and bottom-up. Before presenting that distinction, I should first say something about what distinguishes primary pragmatic processes in general from the more traditional sort of pragmatic process evoked in the Gricean literature: *secondary* pragmatic processes, as I call them.

Secondary pragmatic processes presuppose that something has been said (some proposition expressed). They are inferential processes taking us from what is said, or rather from the speaker's saying of what is

said, to something that (under standard assumptions of rationality and cooperativeness) follows from the fact that the speaker has said what she has said. To the extent that the speaker M-intends the hearer to recognize such consequences as following from her speech act, they form an integral part of what the speaker means by her utterance. That is, roughly, Grice's theory of 'conversational implicature'. An essential aspect of that theory is that the hearer must be able to recognize what is said and to work out the inferential connection between what is said and what is implied by saying it.

In contrast to secondary pragmatic processes, primary pragmatic processes do not presuppose that some proposition has been identified or determined; they are involved in the very determination of what is said. Their input is not a complete proposition, but the linguistic meaning of the sentence, of which the language users need not be consciously aware. (In contrast, as we have just seen, participants in the speech process are aware both of what is said - the input to secondary pragmatic processes - and of what the speaker implies by saying it, as well as of the inferential connection between them.)

I said above that there are two types of primary pragmatic processes. The determination of the reference of indexicals and, more generally, the determination of the content of context-sensitive expressions is a typical *bottom-up* process, i.e. a process triggered (and made obligatory) by a linguistic expression in the sentence itself. For example, if the speaker uses a demonstrative pronoun and says 'She is cute', the hearer must determine who the speaker means by 'she' in order to fix the utterance's truth-conditional content. Similarly, if the speaker uses the genitive construction 'John's car', the heare must determine which relation R is meant to hold between John and the car at issue. The expression itself acts as a variable in need of contextual instantiation; it sets up a slot which the interpreter has to fill. It is in that sense that the pragmatic process at issue here - 'saturation', as I call it - is a bottom-up process. But there are other primary pragmatic processes which are not bottom-up. Far from being triggered by an expression in the sentence, they take place for purely pragmatic reasons.

To give a standard example, suppose someone asks me, at about lunch time, whether I am hungry. I reply: 'I've had a very large breakfast'. In this context, my utterance conversationally implies that I am not hungry. In order to retrieve the implicature, the interpreter must first understand what is said - the input to the secondary pragmatic process responsible for implicature generation. That input is the proposition that the speaker has had a very large breakfast... when? No time is specified in the sentence, which merely describes the posited event

as past. On the other hand, the implicature that the speaker is not hungry could not be derived if the said breakfast was not understood as having taken place on the very day in which the utterance is made. Here we arguably have a case where something (the temporal location of the breakfast event on the day of utterance) is part of the intuitive truth-conditions of the utterance yet does not correspond to anything in the sentence itself ((Sperber and Wilson, 1986, pp. 189-190))[4]. If this is right, then the temporal location of the breakfast event is an *unarticulated constituent* of the statement made by uttering the sentence in that context[5].

Such unarticulated constituents, which are part of the statement made even though they correspond to nothing in the uttered sentence, result from a primary pragmatic process of *free enrichment* - 'free' in the sense of not being linguistically controlled. What triggers the contextual provision of the relevant temporal specification in the above example is not something in the sentence but simply the fact that the utterance is meant as an answer to a question about the speaker's present state of hunger (which state can be causally affected only by a breakfast taken on the same day). While saturation is a bottom-up, linguistically controlled pragmatic process, free enrichment is a top-down, pragmatically controlled pragmatic process. Both types of process are primary since they contribute to shaping the intuitive truth-conditions of the utterance, which intuitive truth-conditions serve as input to secondary pragmatic processes.

According to the view we arrive at, truth-conditional interpretation is pragmatic to a large extent. Various pragmatic processes come into play in the very determination of what is said; not merely saturation - the contextual assignment of values to indexicals and free variables in the logical form of the utterance - but also free enrichment and other processes which are not linguistically triggered but are pragmatic through and through.

8.4 The syncretic view: reconciling tcp with minimalism

Many theorists think one should not allow top-down processes, which are pragmatic through and through, to affect the proposition expressed by an utterance. In order to be part of what is literally said, they claim, a pragmatically provided constituent must at least correspond

---

[4]This is debatable. In (Recanati, 1993, pp. 257-258), I suggest a possible treatment of that example in terms of saturation.

[5]The phrase 'unarticulated constituent' is due to John Perry, who uses is it in a somewhat narrower sense. See Chapter [10] in Perry (1993).

to something in the sentence. It must be 'articulated'. This constraint is what, in previous writings, I referred to as (Pragmatic) Minimalism[6].

Minimalism

What is said is affected by the bottom-up process of saturation but not by top-down processes such as free enrichment.

In those writings I argued against Minimalism, on the following grounds. Once pragmatic interpretation is allowed to play a role in the determination of what is said, it is somewhat arbitrary to set limits to its operation, as Minimalism attempts to do. From a psychological point of view, we cannot separate those aspects of speaker's meaning which fill gaps in the mental representation associated with the sentence as a result of purely semantic interpretation, and those aspects of speaker's meaning which are optional and enrich or otherwise modify the representation in question. They are indissociable, mutually dependent aspects of a single process of pragmatic interpretation (see Recanati (1995) for an illustration of this interdependence).

The suspicion has arisen in several quarters that the quarrel between Truth-conditional pragmatics and Minimalism may well be verbal rather than substantive. To some extent, I agree (see section V below). If the notion of 'what is said' we are trying to characterize is meant to capture the intuitive truth-conditions of an utterance *qua* input to secondary pragmatic processes, then it must be acknowledged that what is said, in that sense, incorporates unarticulated constituents and is therefore affected by free enrichment. (Or so it seems.) But this does not prevent us from defining *another* notion of what is said, conforming to Minimalism. In Recanati (1999) I used subscripts to distinguish the two notions, and I will do so here again. Let 'what is said$_{min}$' be the proposition expressed by an utterance when the effects of top-down pragmatic processes such as free enrichment have been discounted, in accordance with Minimalism; and let 'what is said$_{int}$' correspond to the intuitive truth-conditions of the utterance, which may well result from the operation of such processes. Both what is said$_{min}$ and what is said$_{int}$ are shaped by pragmatic interpretation, but not to the same extent. If I am right what is said$_{int}$ is affected by top-down processes such as free enrichment, whereas the only pragmatic processes that are allowed to affect what is said$_{min}$ are those that are triggered by something in the sentence itself.

---

[6]According to Minimalism, "a pragmatically determined aspect of meaning is part of what is said if and only if its contextual determination is triggered by the grammar, that is, if the sentence itself sets up a slot to be contextually filled" (Recanati, 1993, p. 240).

The view according to which there are two equally legitimate notions of 'what is said', each corresponding to a distinct level in the interpretation of an utterance, I have dubbed the Syncretic View Recanati (2001). It is a reasonable and balanced position. It is actually a mixture of two views: Truth-conditional pragmatics (tcp), and Minimalism or rather a particular version of it (S*-MIN)[7]. TCP consists of theses (i) and (ii), while S*-MIN consists of thesis (iii). The Syncretic View is the conjunction of (i), (ii), and (iii).

TCP

    (i) Pragmatic interpretation is needed to contextually determine the content of context-sensitive expressions. In other words: saturation is a full-fledged pragmatic process.

    (ii) The intuitive truth-conditions of many utterances are also affected by top-down pragmatic processes such as free enrichment.

S*-MIN

    (iii) The proposition literally expressed by the utterance (what is $said_{min}$) is affected by the bottom-up process of saturation but not by top-down processes such as free enrichment.

8.5 Versions of minimalism

As we have just seen, Minimalism per se is not incompatible with TCP. There are several variants of Minimalism, only one of which, so far as I can tell, is actually incompatible with TCP.

Stipulative Minimalism (S-Minimalism) uses the Minimalist constraint as a criterion for demarcating 'what is said'. What is literally said is *defined* as satisfying Minimalism, that is, as being unaffected by top-down factors. Kent Bach ascribes to Grice a version of S-Minimalism. According to Grice's stipulation, Bach says, "what is said must *correspond to* 'the elements of [the sentence], their order, and their syntactic character' (Grice, 1989, p. 87). So if any element of the content of an utterance... does not correspond to any element of the sentence being uttered, it is not part of what is *said*" Bach (2001). Since that is a stipulation concerning the use of the phrase 'what is literally said', there is no way to disagree with such a view, except on terminological grounds. Let me add that, for Grice and most minimalists (though not, as it happens, Bach himself), what is said must be a complete, truth-evaluable proposition. It must be minimal yet truth-evaluable.

---

[7]This stands for 'S*-Minimalism'. See Section 8.5 below for the distinction between that and other forms of Minimalism.

We can make S-Minimalism a little less stipulative by augmenting it with an existence claim. The resulting view, S*-Minimalism, also uses the Minimalist constraint to define 'what is said', but it adds to the definition the following claim: that the notion so defined has a nonempty extension, i.e. that there is a level of meaning that is both minimal and propositional. When S-Minimalism is thus strenghtened into S*-Minimalism, nonterminological disagreement becomes possible. Indeed the view I call Contextualism (Section VI) denies the existence claim which distinguishes S*-Minimalism from S-Minimalism.

Though it conflicts with Contextualism, S*-Minimalism is still compatible with TCP; for the level of meaning it posits, which satisfies the Minimalist constraint by definition, need not be the same level of meaning as that which concerns TCP, namely the intuitive truth-conditions of the utterance (what is said$_{int}$); hence there need be no contradiction between TCP's nonminimalist characterisation of what is said$_{int}$ and S*-Minimalism. (Indeed the Syncretic View incorporates both TCP and S*-Minimalism. It distinguishes between the proposition literally expressed by the sentence, with respect to the context at hand, from what the speaker states in uttering the sentence; what is stated being, in turn, distinguishable from what the speaker merely implies.)

What, then, is the form of Minimalism with which TCP conflicts? As I pointed out above, the quarrel between TCP and Minimalism has been terminological to a large extent. People using 'what is said' in accordance with the Minimalist stipulation have been annoyed when TCP theorists like myself started using the phrase 'what is said' in a different way. Instead of using the Minimalist constraint as a criterion for demarcating what is said, I explicitly put foward a different criterion: the 'availability' criterion, according to which what is said is the proposition determined by the truth-conditional intuitions of the participants in the talk-exchange themselves. But that terminological difference is not the whole story. Once 'what is said' has been demarcated using the availability criterion, it becomes an empirical question whether or not it satisfies the Minimalist constraint. The latter is no longer construed as a defining criterion, but as an empirical characterization. According to TCP, that empirical characterization is falsified by all the cases in which what is said (in the sense of TCP, that is, what is said$_{int}$) involves unarticulated constituents. Thus TCP-theorists claimed to have refuted Minimalism. What TCP-theorists were attempting to refute was neither S-Minimalism nor S*-Minimalism, however, but a third variant which we may call *I-Minimalism* - Minimalism construed as an

empirical characterization of what is said$_{int}$.

I-Minimalism

What is said$_{int}$ is affected by the bottom-up process of saturation but not by top-down processes such as free enrichment.

> I-Minimalism is a nonstipulative version of Minimalism, for the Minimalist constraint is not used as a defining criterion for demarcating what is said. What is said is independently demarcated, using another criterion (the availability criterion), and it is claimed that what is saidint satisfies the Minimalist constraint, as a matter of empirical fact[8].

In a recent paper, Jason Stanley has criticized TCP as empirically inadequate. Whenever an intuitive constituent of what is said seems to be unarticulated, he says, it is in fact articulated by a free variable in logical form. This, he argues, can be established by appealing to the following premiss:

> Since the supposed unarticulated constituent... is not the value of anything in the sentence uttered, there should be no reading of the relevant linguistic constructions in which the unarticulated constituent varies with the values introduced by operators in the sentence uttered. Operators in a sentence only interact with variables in the sentence that lie within their scope. But, if the constituent is unarticulated, it is not the value of any variable in the sentence. Thus, its interpretation cannot be controlled by operators in the sentence (Stanley, 2000, pp. 410-411).

Stanley then uses data of the sort originally collected by Barbara Partee Partee (1989) to show that, in 'each' case in which an alleged unarticulated constituent has been postulated to account for the intuitive meaning of an utterance, one can intuitively 'bind' the alleged unarticulated constituent, i.e. make it vary according to the values introduced by some operator. For instance, the temporal location of the breakfast event, which was said to be an unarticulated constituent of the speaker's response in the example from section III, can be bound by a quantifier. We can say:

---

[8]There are other nonstipulative variants of Minimalism. For example, Kent Bach uses yet another criterion for demarcating what is said. He uses what he calls the IQ test. Whenever we can report an utterance by saying 'The speaker said that . . .', the 'that'-clause expresses what-was-said by the reported utterance. Bach thinks this gives us a test for demarcating what is said. Since the criterion is independent from the Minimalist constraint, whether what is said so demarcated satisfies the constraint is, again, an empirical question. Bach's positive answer to that question therefore gives rise to a fourth variety of Minimalism, which we may call 'IQ-Minimalism'. (I will not discuss it in this paper. Let me simply point out in passing that the IQ test is not equivalent to my 'availability criterion'. Indeed I reject the IQ test, and Bach rejects the availability criterion).

(1) No luck. Each time you offer me lunch, I've had a very large breakfast.

The temporal location of the breakfast event now systematically varies with the temporal values introduced by 'each time you offer me lunch'. It follows (according to the argument) that the alleged unarticulated constituent in the original example was not really unarticulated: it had to be the (contextual) value of a variable in the logical form of the sentence, since without a variable there could not be the sort of binding that occurs in (1).

Clearly, what Stanley is defending in his paper is not S-Minimalism (which needs no defense, since it is vacuously true), nor even S*-Minimalism, but I-Minimalism. It is important to realize that Stanley thus agrees with TCP on the *analysandum*. What he is concerned with, like the TCP-theorist, are the intuitive truth-conditions of utterances[9]. His defense of Minimalism is therefore strikingly unlike the usual defense. The usual defense of Minimalism against TCP consists in arguing that a decent semanticist should be concerned not with 'what is said' in the intuitive sense, but with something more abstract, which satisfies the Minimalist constraint but need not surface to consciousness. That is changing the subject, and we can't help feeling that the TCP-theorist and the Stipulative Minimalist are talking at cross-purposes. But Stanley and the TCP-theorist are clearly talking about the same thing: what is said$_{int}$. They both demarcate what is said using the availability criterion - whether or not there is another legitimate notion of what is said. They agree that in the breakfast example, the temporal location of the breakfast is a constituent of what is said in that sense. Yet Stanley contends that what is said, in that sense, conforms to Minimalism as a matter of empirical fact, while the TCP theorist contends, also on empirical grounds, that it does not. The disagreement here is genuine empirical disagreement (see (Stanley (2000)) and (Recanati (2002)) for the details of the debate).

8.6 From Radical Literalism to Contextualism

Let us take stock. From what I have said it follows that there are four basic positions concerning the role of context in the determination of truth-conditions. *Radical Literalism* holds that the truth-conditions of a sentence are fixed by the rules of the language quite independent of the users's beliefs and intentions[10]. TCP rejects that view on two grounds:

---

[9] "Accounting for our ordinary judgments about the truth-conditions of various sentences is the central aim of semantics" (Stanley and Szabo, 2000, p. 240).

[10] For a recent defense of radical literalism, see Borg (2005).

(i) semantic underdetermination makes it unavoidable to appeal to speaker's meaning in determining truth-conditions; (ii) speaker's meaning also comes into the picture via the provision of unarticulated constituents which enrich the truth-conditions without corresponding to anything in the sentence itself. There is an intermediate position between *Radical Literalism* and TCP, however; for it is clearly possible to accept TCP's first claim, pertaining to semantic uderdetermination, without accepting the second claim, pertaining to unarticulated constituents. The intermediate position, which I ascribed to Stanley, may be called Moderate Literalism. According to Moderate Literalism, we need to appeal to speaker's meaning in determining truth-conditions, but we do so only when the sentence itself demands it.

TCP, as we have seen, is compatible with S*-Minimalism: the view that there is a level of meaning which is both (i) propositional (truth-evaluable) and (ii) minimalist, i.e. unaffected by top-down factors. TCP says that the intuitive truth-conditions of an utterance *are* affected by top-down factors and incorporate unarticulated constituents, but this is compatible with the existence of *another* level of meaning from which such constituents are banned. The conjunction of TCP and S*-Minimalism is the *Syncretic View* (section IV). It claims that there are two notions of what is said: what is strictly and literally said (what is said$_{min}$) is minimalist, but what is said in the intuitive sense (what is said$_{int}$) is not. On that view what is said in the intuitive sense is still distinct from what is merely 'implied' (in, again, the intuitive sense).

Though it is compatible with S*-Minimalism, TCP does not entail it. A more radical version of TCP actually rejects S*-Minimalism. According to that version, which I call *Contextualism*, 'no proposition could be expressed without some unarticulated constituent being contextually provided' (Recanati, 1993, p. 260), hence the extension of the notion of 'what is said' defined according to the Minimalist stipulation is empty: there is no level of meaning which is both (i) propositional and (ii) minimalist. Such a radical view has been defended by Charles Travis, John Searle, and a few others. Searle, for example, argues that a determinate proposition can be expressed only against a background of unarticulated assumptions. He gives the following example of unarticulated assumptions at work in understanding a simple utterance:

> Suppose I go into the restaurant and order a meal. Suppose I say, speaking literally, 'Bring me a steak with fried potatoes.' (...) I take it for granted that they will not deliver the meal to my house, or to my place of work. I take it for granted that the steak will not be encased in concrete, or petrified. It will not be stuffed into my pockets or spread over my head. But none of these assumptions was made explicit in the

literal utterance. (Searle, 1992, p. 180)

Though unarticulated, those assumptions contribute to determining the intuitive conditions of satisfaction (obedience-conditions, truth-conditions, etc.) of the utterance. The order 'Bring me a steak with fried potatoes' does not count as satisfied if the steak is delivered, encased in concrete, to the customer's house. It is mutually manifest to both the hearer and the speaker that the speaker intends the ordered meal to be placed in front of him on the restaurant table he is sitting at, etc. Though not explicitly said, that is clearly part of what is meant. Yet one does not want to say that that aspect of utterance meaning is conveyed indirectly or nonliterally (as when one says something and means something else). The utterance 'Bring me a steak with fried potatoes' is fully literal. It is a property of literal and serious utterances that their conditions of satisfaction systematically depend upon unstated background assumptions.

In that sort of case a Syncretist may be willing to insist that what is strictly and literally said is free from the relevant assumptions: the order 'Bring me a steak with fried potatoes' would be literally satisfied if the steak was delivered, encased in concrete, to the customer's house. That is precisely what the Contextualist denies. Another example given by Searle will help to make that point. The word 'cut' is not ambiguous, Searle says, yet it makes quite different contributions to the truth-conditions of the utterance in 'Bill cut the grass' and 'Sally cut the cake'. That is because background assumptions play a role in fixing satisfaction-conditions, and different background assumptions underlie the use of 'cut' in connection with grass and cakes respectively. We assume that grass is cut in a certain way, and cakes in another way. The assumed way of cutting finds its way into the utterance's truth-conditions:

> Though the occurrence of the word "cut" is literal in [both] utterances..., and though the word is not ambiguous, it determines different sets of truth conditions for the different sentences. The sort of thing that constitutes cutting the grass is quite different from, e.g., the sort of thing that constitutes cutting a cake. One way to see this is to imagine what constitutes obeying the order to cut something. If someone tells me to cut the grass and I rush out and stab it with a knife, or if I am ordered to cut the cake and I run over it with a lawnmower, in each case I will have failed to obey the order. That is not what the speaker meant by his literal and serious utterance of the sentence. (Searle, 1980, pp. 222-223)

Now the Syncretist will assume that a sentence such as 'Cut the grass' expresses something that has literal conditions of satisfaction

quite independent of any background assumption; something very abstract, involving the constant, underspecified meaning of 'cut' and not the definite sense it takes on particular uses (or types of use). Stabbing the grass with a knife and running over it with a lawnmower are two ways of literally obeying the order 'Cut the grass', on this view. But the Contextualist stands skeptical. To get something genuinely truth-evaluable, he holds, we need background assumptions. We cannot specify a determinate proposition which the sentence can be said literally to express, without building unarticulated assumptions into that proposition. The best we can do is to construct a disjunction of the propositions which could be determinately expressed by that sentence against alternative background assumptions.

To sum up, the four positions in the ballpark can be characterized in terms of their answers to three basic questions (see Table below).

	Q1. Semantic under-determination?	Q2. Unarticulated constituents?	Q3. Minimal proposition?
Radical Literalism	no	no	yes
Moderate Literalism	yes	no	yes
Syncretic View	yes	yes	yes
Contextualism	yes	yes	no

The three basic questions can be spelled out as follows:

Q1. *Do we have to appeal to speaker's meaning to fix truth-conditions?*
Radical Literalism makes a negative answer. That is what renders that position hopeless: it ignores the phenomenon of semantic underdetermination, which is characteristic of natural language. All three other positions make an affirmative answer to that question.

Q2. *Is there free enrichment of truth-conditional content? Are there 'unarticulated constituents'?*
All versions of Literalism deny that there is such a phenomenon. Literalism is the view that the truth-conditional content of an utterance is determined by the linguistic material in the uttered sentence. No contextual influences are allowed to affect truth-conditional content unless the sentence itself demands it. As Stanley puts it, "all effects of extra-

linguistic context on the truth-conditions of an assertion are traceable to elements in the actual syntactic structure of the sentence uttered" (Stanley 2000: 391). In contrast, all versions of TCP acknowledge the phenomenon of free enrichment and depart from Literalism by opting for a nonminimalist construal of what is said in the intuitive sense.

Q3. *Is there a level of meaning that is both propositional and minimal?*

That level of meaning need not be what is said in the intuitive sense, hence it is possible to opt for a nonminimalist construal of what is said$_{int}$, as TCP does, while making an affirmative answer to question 3. This corresponds to the Syncretic View - a moderate version of TCP. The more radical version, Contextualism, answers question 3 negatively.

I conclude that there are two interesting debates in the area surveyed in this paper. One is the debate between TCP and Moderate Literalism over unarticulated constituents. TCP says that there are unarticulated constituents: the intuitive truth-conditional content of utterances is affected by free enrichment. Moderate Literalists deny this. The other interesting debate is that over minimal propositions, and it goes deep. It is commonly assumed that *sentences* have truth-conditions ('with respect to contexts', if the language contains context-sensitive elements). According to Contextualism, however, it is not natural language sentences, not even sentences 'with respect to context', which have truth conditions, but full-blooded speech acts - meaningful actions performed by rational agents. Sentences have truth-conditions only derivatively[11]

Acknoledgments

I am indebted to Varol Akman and Andrea Bonomi for their useful comments on an earlier version of this chapter.

References

Austin, J. 1975. *How to Do Things with Words*. Clarendon Press, Oxford.

Bach, K. 2001. You don't say. *Synthese* 128:15–44.

Barwise, J. and J. Perry. 1983. *Situations and Attitudes*. Cambridge, MA: MIT Press.

Borg, Emma. 2005. Saying what you mean: Unarticulated constituents and communication. In R. E. . R. J. Stainton, ed., *Ellipsis and Non-sentential Speech*, pages 237–262. Dordrecht: Kluwer Academic Publishers.

---

[11]Historically, Contextualism is associated with Wittgenstein and ordinary-language philosophers. It is widely considered to have been refuted by e.g. Grice, Geach, and others, but I do not think that is right: Contextualism, I take it, still is a live option. Of the four positions I mentioned, only Radical Literalism is (in my view) a nonstarter.

Fodor, J.A. 1983. *The Modularity of Mind*. Cambridge, Mass.: MIT Press.

Grice, H.P. 1989. *Studies in the Way of Words*. Harvard University Press, Cambridge, Massachusetts.

J.Stainton, Robert. 2005. In defense of non-sentential assertion. In Z. G. Szabó, ed., *Semantics vs. Pragmatics*, pages 383–457. Oxford: Oxford University Press, Oxford.

Kaplan, D. 1989. Demonstratives. In J. Almog, J. Perry, and H. Wettstein, eds., *Themes from Kaplan*, pages 481–563. Oxford: Oxford University Press.

Partee, B. 1989. Binding implicit variables in quantified contexts. In *Proceedings of the Chicago Linguistics Society*, vol. 25, pages 342–365.

Perry, J. 1993. *The Problem of the Essential Indexical and Other Essays*. Oxford University Press, New York. Enlarged edition, CSLI Publications, Stanford, 2000.

Recanati, F. 1993. *Direct Reference: from Language to Thought*. Blackwell, Oxford.

Recanati, F. 1995. The alleged priority of literal interpretation. *Cognitive Science* 19:207–232.

Recanati, F. 1999. Situations and the structure of content. In K. Murasugi and R. Stainton, eds., *Philosophy and Linguistics (Boulder, CO, USA)*, pages 113–165.

Recanati, F. 2001. What is said. *Synthese* 128:75–91.

Recanati, F. 2002. Unarticulated constituents. *Linguistics and Philosophy* 25:299–345.

Searle, J. 1980. The background of meaning. In J. Searle, F. Kiefer, and M. Bierwisch, eds., *Speech Act Theory and Pragmatics*, pages 221–232. Dordrecht: Reidel.

Searle, J. 1992. *The Rediscovery of the Mind*. MIT Press.

Sperber, D. and D. Wilson. 1986. *Relevance: Communication and Cognition*. Basil Blackwell.

Stanley, J. 2000. Context and logical form. *Linguistics and Philosophy* 23:391–434.

Stanley, J. and Z. Szabo. 2000. On quantifier domain restriction. *Mind and Language* 15:219–261.

Travis, C. 1981. *The True and the False: the Domain of the Pragmatic*. J. Benjamins, Amsterdam.

Travis, C. 1985. On what is strictly speaking true. *Canadian Journal of Philosophy* 15:187–229.

9

Context and Contract

CARLO PENCO

9.1   Introduction

The main point of this paper is the claim that a strong notion of cognitive context can answer the needs of a representation of dialogue context, with a higher generality than the "normative" notion suggested by Gauker. I will discuss some well known claims in the literature about communication and context, and I will suggest giving a central role to the notion of contract or semantic bargaining and to the normative constraints of indexicals and anaphora.

In (§ 9.2) I will classify different concepts of context and will define a setting for treating discourse context inside a general framework of cognitive context. I will then contrast the claims of Stalnaker and Gauker on the concept of discourse context, showing some aspects of convergence between the two proposals on the idea of normativity (§ 9.3). I will then give an account of a discussion in an example given by Gauker to support the idea of normative context (§ 9.4).

§ 9.2, 9.3, 9.4 are the background for the main discussion in § 9.5. Here I will try to show some limitations of Gauker's concept of normative context. I will claim that a notion of cognitive context may help us to revise our ideas of where to place normativity in discourse context, when semantic negotiation is at stake[1].

9.2   Types of context

"I am here now" is different from "7+1=8", because the meaning of the first sentence is context dependent, while the second is not. The normal way to explain that is the following: to interpret the first sentence you need to know who the speaker is, while to interpret the second one, you do not need to know who is speaking. This is misleading, because we may say that the meaning of "7+1=8" in the theory of natural numbers is different from the meaning "7+1=8" in the theory of integers. We have different rules and different algebraic structures and only in the second case may we infer "therefore 7-8 = -1", which is not a valid formula in the theory of natural numbers. Hence, the meaning of "7+1=8" is also context dependent.

We have here a basic contrast between two ways of understanding "context" and "context dependence": *objective* context, or the state of affairs (location, time, speaker) in which a sentence is uttered, and *cognitive* context, or the theory in which a sentence is interpreted. Modern formal logic was invented for treating mathematical theories and mathematical reasoning, which is relatively independent from objective context – from speaker, time and location of the utterance. In the '70s, attention to the *objective* context was developed from the effort to widen the scope of logic, in order also to treat utterances of natural language and to disambiguate different aspects of meaning. Since then it has become common to distinguish different levels of meaning, as widely discussed by Akman 2006 on the threefold distinction made by Strawson among linguistic meaning, reference and contextual or illocutionary meaning. Recanati 2001 gives the "standard" semantic view as a distinction of (i) linguistic meaning, (ii) what is said, and (iii) what is implicated. In an analogous way, Perry 1998 distinguishes pre-semantic, semantic and post-semantic uses of context, which define linguistic or literal meaning, reference and presupposed meaning.

"Context" is not a natural kind term, but a term for a concept we have invented to understand the workings of our language and knowledge. We have to distinguish, in the relevant literature, different ways to treat the term "context". We will refer here to three different notions of context, to check whether they are independent or reducible to one another at least methodologically[2]. Here are the three main general

---

of meaning negotiation.

[2] An *ontological* reduction would mean a choice between realism and antirealism: for instance a reduction of objective context to cognitive context would amount to an extreme anti-realist ontological stance, very near to an idealist position. It would be like to claim that what exists is just reducible to our interpretation of what exists. A methodological reduction is a more modest and reasonable strategy,

concepts:

(1) Metaphysical or Objective or Semantic Context: the state of affairs in the actual world, relative to an utterance, or context of the utterance. We might give a formalization of such a concept with a tuple of parameters ⟨speaker, time, place,...⟩. Kaplan 1978, 1989 is one of the best advocates of this kind of concept, which is fit to be embedded in a model-theoretic semantics, and represents the metaphysical state of affairs - what there is. Kaplan's treatment of objective context implies a distinction of at least two levels of meaning[3], character and content (briefly, linguistic meaning and truth conditional meaning) enlarging the traditional analysis of mathematical logic. Lewis 1980[4] proposes to enrich objective context with other situational elements such as presuppositions and standard of precision. Another way of speaking of objective context is Perry's *semantic context*, to be identified with what is needed to give an evaluation to indexical expressions, after disambiguating the literal or linguistic meaning of the words (Perry (1998)). Bach 1996 speak of "restricted" or narrow context, referring to the variables used to fix the evaluation of indexicals (speaker, time, place), contrasting it with a wider notion of context, related to any contextual information relevant to determining the speaker's intention. An analogous distinction is given by Gauker 1998b, who uses a different terminology and distinguishes *situational* aspects (which correspond to what we have here called "objective" context) from *propositions* that should be presupposed in view of the aim of a conversation (which correspond to what *he* calls "objective context"). To avoid misunderstanding, I will use the term "normative context" for the latter concept, and use the term "objective context" to refer to the situational aspects of the context of utterance.

(2) Cognitive or Subjective or Pragmatic Context: a point of view of a situation, or a theory in which a situation can be considered or described. This notion is apt for distinguishing the different meaning of

---

which aims to show that one representation may be preferred to another for certain reasons and purposes.

[3]Kaplan gives some hints regarding aspects of the Fregean theory, which have not been considered in logic, giving his own alternative. Referring to different aspects of the Fregean notion of sense, Perry 1977 makes an analogous bipartite distinction, which will be developed further in his work. These distinctions are compatible with the "standard view" in semantics (as presented by Recanati (2001)) as a three-layer distinction of linguistic meaning, what is said, what is implicated.

[4]The technical developments of these ideas are linked to the proposal of using a double indexing, one for context and the other for time and possible world. The idea was originally Kamp's and later developed by Kaplan. Lewis remarks that the need for a double indexing must be kept for features which can be *shifted*, that is time, location, possible world and standard of precision

some mathematical or logical formula depending on the theory in which the formula is used (think of the difference between a classical or intuitionistic interpretation of a formula). Aspects of cognitive context have often been included in the objective context, which may be thought of as also including mental states or beliefs of the speakers. This seems to be consonant with Lewis's definition of context and with Bach's concept of wide context. A cognitive characterization stresses the *theory-laden* aspect of the cognitive context. From this viewpoint, cognitive context can be represented as a local theory consisting of ⟨Language, Axioms, Rules⟩. Recently this idea has been developed by McCarthy in 1993 and Giunchiglia in 1993, with the basic motivation that (i) for every axiom we may find, with some ingenuity, a more general context where the precise form of the axiom does not hold (the problem of generality); and (ii) for any situation, we need to use the smallest possible amount of information to reasonably treat any problem avoiding combinatorial explosion (the problem of locality). Any interpretation and evaluation of utterances needs a defined cognitive context in which the utterances receive both meaning and semantic value. Benerecetti 2006 show how this contextualist stance includes reference to external parameters of a context.

How is this different from objective context? Partly the difference is that objective context aims to give the objective features necessary to evaluate an utterance, while cognitive context, as a theory about a situation, typically aims to give a defeasible point of view (of individuals, groups, institutions, databases...) about a situation. Perry's *post-semantic context* can be considered a kind of cognitive context, because it represents what is needed to interpret a sentence even when the semantic evaluation of the indexicals has already been given. Perry's idea implies that we need to first have the evaluation of indexicals, and eventually, in some peculiar cases, a further evaluation. It has been suggested by Bianchi 1999 that this kind of post-semantic context could be intended as an instance of a more general pragmatic context dependence, valid for every utterance, as in the radical contextualism put forward by Travis 1997 and Searle. This point of view is developed by Recanati (this volume) who reverses the order of "standard" explanation of levels of meaning: first the context, then the meaning. Recanati speaks of pragmatic context (stressing the reference to actions or speech acts); however, the direction of his work seems compatible with a treatment of context as characterized as a (local) theory of a situation.

(3) Discourse or Dialogue or Conversational Context: to describe a conversation you need to refer to (i) an objective context of utter-

ance, that is speakers, location and time of the conversation, (ii) a representation of the different cognitive points of view or background assumptions of the interlocutors. Therefore, at first sight, it seems that treating discourse context requires both objective and cognitive context. Discourse Representation Theory (DRT) began by developing a theory of discourse context, in a way that made it relative to the processing of discourse from the standpoint of a hearer. In this case, the representational structure of the elements of discourse works as the context in which to interpret a new sentence following in the discourse[5]. In his development of DRT, Kamp (1993), even though he focuses strongly on syntactic aspects, accepts the need for providing a model theoretical semantics. Stalnaker 1999 seems to follow the opposite path: he begins with a possible-worlds semantics to recognize the need and the importance of syntactic structure. He gives therefore a double representation of discourse context consisting of shared information about (i) the subject matter of the discourse, represented in a possible-worlds semantics (ii) facts about the discourse, including syntactic aspects, which need to be taken into account (e.g. the specific language in which it is produced). Still, we need to discuss which conceptual tools to use (shared presupposition, metaphysical context, normative context, cognitive contexts, semantic or syntactic representations?), depending on the kind of problem we face in treating discourse context.

In Penco (1999, 2000) I tried to show the plausibility of the reduction of theories about metaphysical context to theories about cognitive context[6]. The reduction is useful to mark the fact that what is "objective" is always described in a perspective, and the idea of objectivity is derived from our disagreement about what we claim is truly so. Any description purports to represent objective reality, and at the same time it is given inside a point of view, which can always be revealed as mistaken or epistemically constrained. Any evaluation of the actual world in which an utterance is made is dependent on the cognitive access of the speakers. We may disagree on the evaluations to be given even regarding speaker, time and location, or even be deprived of access to the evaluation of some instance of either expressed or "unarticulated" indexicals.

This last claim needs clarification; when Perry 1998 claims that there is an unarticulated component in a sentence like "it rains", he seems to imply that to evaluate "it rains" it is mandatory to fill the variable

---

[5]Some more detailed comments about the DRT are given in Arló-Costa (2006).

[6]Benerecetti et al. (2003) gives a suggestion on a formal reduction of Kaplan's logic of demonstratives inside a cognitive-context framework based on a multi-context Logic.

for the place, which is not expressed in the sentence. But we might have a case where it is neither necessary nor welcome to evaluate the place: Recanati[7] gives an example of a situation where we have access to information that it rains somewhere, while not having access to the place where it is raining. The example shows that information about the place is not always relevant for the evaluation of an utterance of "it rains". We may say that from a metaphysical point of view there is a place where it is raining, and from an epistemic point of view we have no way to access it. In fact, even if we assume that a place where it rains exists, the individuation of the place is irrelevant for the evaluation. The point of the example is similar to the above-mentioned point given by McCarthy 1987 about the "generality" constraint: we may always find a (cognitive) context where we cannot evaluate a sentence in the "intended" way. The interpretation depends then on the number of variables or parameters we decide or need to consider. Therefore, we always have to evaluate a sentence inside a theory that expresses what is needed, to make its interpretation relative to the theory (e.g. a theory which asks us not to evaluate the parameter "location", when this is not accessible or relevant, or not to evaluate the parameter "time" when this is irrelevant). We should therefore at least require the following:

1. The evaluation of a sentence depends on pragmatic parameters, including speaker, place and time. However the evaluation of such parameters is not always accessible to the speakers themselves (think of "here" and "now" without explicit or external knowledge of time and place, or uses of "I" in cases of amnesia about personal identity).

---

[7]Recanati (2002) says: "Can we not imagine a context in which 'It is raining' would be evaluable even if no particular place were contextually singled out? I have no difficulty imagining such a context. I can imagine a situation in which rain has become extremely rare and important, and rain detectors have been disposed all over the territory (whatever the territory - possibly the whole Earth). In the imagined scenario, each detector triggers an alarm bell in the Monitoring Room when it detects rain. There is a single bell; the location of the triggering detector is indicated by a light on a board in the Monitoring Room. After weeks of total drought, the bell eventually rings in the Monitoring Room. Hearing it, the weatherman on duty in the adjacent room shouts: 'It's raining!' His utterance is true, iff it is raining (at the time of utterance) in some place or other. The fact that one can imagine an utterance of 'It's raining' that is true iff it is raining (at the time of utterance) in some place or other arguably establishes the pragmatic nature of the felt necessity to single out a particular place, in the contexts in which such a necessity is indeed felt. If that is right, there is no need to posit a lexically specified argument-role for a location in the sub-atomic structure of the verb 'rain': 'Rain' is like 'dance' and other action verbs (...). That raining must take place somewhere or other is a metaphysical fact, not a linguistic fact. That fact does not prevent an utterance like ["It is raining"] from expressing a fully determinate proposition even if no place is contextually provided."

2. The value of these parameters must be represented - when possible - as a part of the cognitive state of theory of a particular agent: speaker, hearer, reporter, interpreter. The role of interpreter is so basic as to be assumed without taking notice of it: the interpreter (sometimes identified with the canonical observer) is what is normally called "us" or "we" in philosophical papers. Here it is normally assumed that we (writers and readers) know the truth.

3. Different agents may give different interpretations to these parameters; therefore we need to have representations of (i) the cognitive contexts in which the evaluation of the parameters is made and (ii) the relations among these different points of view.

What is the case when *discourse or conversational context* is at stake? We need to give the right place to cognitive contexts or theories of speakers, hearers and observers. What is mainly relevant for us is a good representation of the interplay of cognitive contexts, which should account for different semantic evaluations among speakers (and sometimes it might be or sometimes it might *not* be possible for *us*, the observers, to give an assessment of these differences).

The claim advanced here is that only this work of representing the interaction of different epistemic or cognitive contexts can provide the ground for a plausible theory of communication that lies behind any analysis of discourse context. This claim has to be compared with other alternative theses; in the next paragraph I will then discuss two alternative claims on communication and context, presented by Stalnaker and Gauker.

9.3   Communication, Normativity, Misunderstandings

Following the ideas of Grice, Stalnaker 1999 criticizes the linguistic turn in philosophy and insists on the priority of intention over language, as if language were only a means by which to express a previously defined intention in our minds. What then is communication? According to the Grice-Stalnaker stance, communication is the successful passage of a proposition a speaker intends to convey to the hearer, and the recognition on the part of the hearer of the proposition uttered by the speaker and of the intention of the speaker, so that the hearer understands the proposition. Given this definition, a definition of context follows. Discourse context (assuming that discourse aims at successful communication) is

> the set of presuppositions a speaker holds to be common assumptions
> with other interlocutors in the dialogue.

This definition sounds very similar to what Davidson (1986) calls "prior theory", that is, the theory the speaker believes the hearer has in mind before beginning a dialogue. There are two main differences with Davidson: (i) Stalnaker does not speak in term of a Davidsonian theory of meaning, but of the discourse context as a more abstract representation in term of possible worlds; (ii) he claims that, at least in the "normal" case, we have to posit a set of shared assumptions. A speech act, typically an assertion, makes a change in this set, making people enlarge it with new information[8].

These ideas have been contested by Gauker 1997 who reacts to the above-sketched definitions of context and communication, treating them as examples of the "expressivist theory of communication" (ETC). According to ETC "the primary function of language is to enable speakers to convey propositions to hearers" (1997, p.5). The alternative view claims that communication, which uses language essentially, is a "matter of getting people to do things in the course of mainly cooperative interactions" (Gauker (1994), p.3-4). From this viewpoint communication is not grounded on sharing propositions or on detecting what intentions people have in mind, in a language-independent manner. Consequently, discourse context cannot be defined simply as a set of possible worlds presupposed in a dialogue. It would be both irrelevant to the goal of

the conversation and computationally intractable[9]. The set of propositions, which constitutes discourse context, is not a set of shared assumptions, but

> the set of propositions, which should be considered for attaining the goals of the conversation.

In these claims we have an apparent radical alternative. The main contrast appears to be the role given to normativity in describing discourse context. Gauker suggests that sharing propositions is not a prerequisite of communication. Context is defined by which propositions *should* be taken for granted when the speaker chooses her words and by which propositions an interlocutor *ought* to acknowledge for achieving the goals of the conversation. In short: propositions are normatively shared and descriptively not shared; they are not what is presupposed in conversation, but what *should* be presupposed for the specific goal of the discourse. I am not sure how deep the contrast between Stal-

---

[8] Apparently the updating of beliefs does not follow this simple pattern; but on this see the discussion by Arló-Costa (2006) on the "persistent theory of belief."

[9] Arló-Costa (2006) suggests that a theory of the kind of Stalnaker's is "too big" for a descriptive purpose. See also analogous remarks in Penco (1999)

naker and Gauker goes on this point. Actually Stalnaker 1999 (p.10) also asserts that the speaker *should* presuppose that the hearers "have whatever information is required to interpret what he is saying". Therefore, the body of information, which is supposed to be shared, is normally intended as the information that *should* be shared. Assuming - as Stalnaker does - that the speaker should presuppose that the hearers have the information "required" is a normative step which goes in the direction of the "objective" (normative) context as given by Gauker.

Putting aside the problem of the relative priority of language and intentions (where there seems to be a real disagreement between Stalnaker and Gauker), in this paper I will focus on the role of normativity which is explicitly developed in Gauker and just hinted at in Stalnaker. Both theories (but I will discuss mainly Gauker's) have to face the problem of the explanation and clarification of mistakes and misunderstandings in communication. In Stalnaker's framework, the problem is placed in the discussion of "abnormal" cases of communication where there is no shared assumption, but there are different points of view among different speakers. The analysis of abnormal situations implies the analysis of how people arrive at a common interpretation of what is said. A set of shared propositions, which is *not* presupposed in "abnormal" cases, becomes an *aim* of the discourse; it is what speakers *converge* towards in a conversation, from the point of view of Davidson's framework. On the contrary, Gauker's normative assumption implicitly suggests avoiding irrelevant analysis of the problem of convergence of belief sets (problem to be left to empirical or psychological research). From his standpoint, each dialogue has a set of propositions normatively ("objectively") given as presupposition. A central notion in Gauker's framework is the notion of "goal of the conversation," which gives the main motivation for defining normative context. In this setting we should be obliged to consider all misunderstandings as grounded on mistakes regarding the objective context constituted by the "norm" of the conversation given by its goal. But are misunderstandings always to be considered as based on "mistakes"? And if they are derived from mistakes, are these mistakes always relative to an objective norm or set of propositions, or may they be mistakes regarding possible equiprobable interpretations? Up to which point are we bound to assume a context - intended as a set of presuppositions or information - as "normative"?

In order to answer this kind of question, I will present a debate between Gauker and Van Deemter on the value of the expressivist view of communication, and I will analyze the debate with respect to the problem of semantic negotiation or semantic bargaining, as already suggested in Sbisà (1999). Even if I strongly appreciate the idea of

normative context à la Gauker, I will try to show some limitations of his normative concept of context when semantic negotiation is concerned. Assuming that - at least in some cases - we cannot properly speak of "normative context", I will try to find other places for normativity to support the idea of semantic bargaining.

Last, but not least, we need to be clear about the notion of sharing. First of all there is a distinction between social sharing and individual sharing. *Social sharing* is sharing among everybody; it is collective and distibuted, like sharing different aspects of meaning in a social division of labour. The meaning is shared by the community, but not every individual belonging to the community possesses all the aspects of meaning, and most people defer relevant aspects to experts. *Individual sharing* is sharing entertained by *each* individual, in a situation in which people are said to have the *same* information or background. In this case we have to distinguish three aspects: subjective sharing, objective sharing, normative sharing. *Subjective* sharing implies awareness, and happens when every participant to the conversation is aware of what the other individuals presuppose. This is highly implausible, and may be stated only as a first approximation at a very general level of stereotypes. *Objective* sharing is a descriptive fact about a dialogue, where we may assert that all participants share the same information, even if they are not aware of that. *Normative* sharing is linked to the information people should have for the correct development of the conversation. Apparently, in an idealized situation, what is descriptively shared (objective sharing) collapses in what is normatively shared (normative sharing). We need to distinguish among these different aspects of sharing, and to clarify which level of information is shared in which ways. I will insist upon the idea of an objective sharing of strategic rules for managing anaphora and indexicals. This sharing of strategic rules is needed for retrieving the different presuppositions of different speakers in a dialogue.

9.4 Domains of discourse and the Expressive Theory

The example (Gauker (1997)): Tommy meets Suzy who has some white and red marbles on the floor in front of her. Tommy says "all the red ones are mine!" and Suzy answers: "no, they are not"[10]. Actually

---

[10]Susy could also say: "No, that is not true". "No they are not" stresses the difference on the domain of discourse; "No, that is not true" stresses the different evaluation of the proposition. The problem in the two cases is: (1) *which* marbles? (2) *which* proposition? I will discuss here the first case, which deals with ambiguity inside restricted quantification. I will discuss more explicitly the second case at the end of the paper.

(we know that) Tommy means the red marbles in his room, and Suzy means the red marbles in her room on the floor in front of her. Gauker suggests that, within the expressivist theory of communication (ETC), it is impossible to give a semantic evaluation to what they are saying because it is impossible to decide which proposition they are discussing. ETC cannot use any of the three possibilities open for interpreting the sentence uttered by Tommy (true, false, neutral):

- ETC cannot accept that the proposition uttered by Tommy is false, because, following Grice in giving priority to the speaker's meaning, the hearer's objective should be to recognize the proposition the speaker intends the hearer to recognize.
- ETC - at least in its contextualist version - cannot accept that the proposition is true, because the setting of the situation suggests that the "intended" interpretation should pick up the marbles in front of Suzy.
- ETC cannot accept the neutral interpretation (there is no domain of discourse or there are two domains) because in this case we could not recognize the different attitudes of Tommy and Suzy toward the same utterance.

An attempt to defend the expressive view of communication, and to show that language is used to enable speakers to convey propositions to hearers, is given by Van Deemter, who uses the tools of underspecified logical languages for that purpose. I use his position as an example of one of the different possible tools aimed at giving space to aspects not considered by Gauker. Local model semantics as given in Ghidini and Giunchiglia (2001) might give analogous results; I refer to Van Deemter because of his particularly clear presentation of the example in question. In underspecified logical languages (Van-Deemter and Peters (1996)) the semantic interpretation attributes a set of semantic values to ambiguous expressions: to give different interpretations to the same expression we may use an interpretive mode $m_m = \langle D, I \rangle$, where the interpretation function maps *occurrences* of predicates into subsets of the domain. Skipping details which can be found in Van Deemter's papers, let me come to the point. The context ambiguity is treated as lexical ambiguity, using an ambiguous constant DoD ("Domain of Discourse") which selects the relevant interpretation of the occurrence of Tommy's utterance. Depending on two different modes - Tommy's mode$_\mathbf{t}$ or Suzy's mode$_\mathbf{s}$ - the sentence

$$T = \forall\ x \in D\ (\text{RedMarble}\ (x) \longrightarrow \text{Tommy's}\ (x))$$

receives two different interpretations (the classical interpretation here attributes two different subsets of the domain D to the non-ambiguous constants $\text{DoD}_\mathbf{t}$ and $\text{DoD}_\mathbf{s}$):

$$T_\mathbf{t} = \forall\, x \in \text{DoD}_\mathbf{t}\ (\text{RedMarble (x)} \longrightarrow \text{Tommy's (x)})$$
$$T_\mathbf{s} = \forall\, x \in \text{DoD}_\mathbf{s}\ (\text{RedMarble (x)} \longrightarrow \text{Tommy's (x)})$$

$T_\mathbf{t}$ and $T_\mathbf{s}$ express different propositions, depending on the two different Domains of Discourse. However, the interpretation of the negation of the sentence T (uttered by Suzy) is a function of the interpretation of T. Under $\text{mode}_\mathbf{t}$ T is true and its negation false; under the $\text{mode}_\mathbf{s}$ T is false and its negation true. In the course of conversation, the interlocutors have to realize the so-called "lexical" ambiguity of the term "red marbles": they have to realize that they are giving different semantic interpretations to the same expression. This is a fairly good example of a neutral perspective, a way to make explicit the idea of "conveying a proposition" suggested by the expressive theory of communication:

A speaker α conveys a proposition p to the hearer β
iff α utters U, β hears U and $m_\beta(U) = p$

Given this framework, Van Deemter discusses the problem of the resolution of the conflict. He resolves it rapidly, considering the follow-up to the dialogue as a prototypical case where eventually the proposition the speaker wanted to convey is recognized and the proposition the hearer erroneously (but with some reason) took him to express becomes obsolete. Here is the first part of the imagined dialogue:

T. All the red ones are mine!
S. No they are not!
T. Yes, they are. Mom gave them to me.
S. Dad gave me those marbles and Mom doesn't even know it.

At this point the dialogue has a natural break. Gauker (1998b) (p.27) hints at two possible developments; either Tommy explicitly says that he is not talking about Suzy's marbles, or Suzy might ask *which* marbles Tommy is talking about. For simplicity's sake, let us follow Van Deemter in developing the first suggestion (but Gauker's point that there is a break in conversation here which might be taken either by Tommy or Suzy is relevant, and we will come back to this later). Here is a possible second part of the dialogue:

T. I'm not talking about those marbles; I'm talking about the marbles in *my* room.
S. Oh. I see! You are referring to the marbles in your room; then sure, they belong to you.

Semantic barganing is used here by Van Deemter to support evidence in favor of an expressivist theory of communication. What is Gauker's answer? Gauker reacts saying that a discourse, if it is to be understandable, should have a unique interpretation, therefore a unique domain of discourse, which is the one (normatively) relevant. In Van Deemter's perspective, on the other hand, if we are expressivists, we should have a unique domain of discourse only when the two domains of the interlocutors coincide. In this way, Gauker answers, to have an opinion on the domain of discourse, expressivists need to go through what the speakers take to be the domain, in an endless recursion of subjective perspectives, perhaps until they find a casual, fortuitous, coincidence. When Suzy says that what Tommy says is false, we may claim that the circumstances were misleading and Suzy was justified in asserting that. But our claim presupposes that we can give an account of what the domain of discourse really is, besides what Tommy and Suzy have "in mind" (Gauker (1998a), p.450).

I am not so sure that I am an expressivist (probably I am, at a certain degree). However my point will not be to decide what an expressivist has to do, but simply to answer the following question: is the only alternative to normative context a theory that has to reduce discourse context to a causal coincidence of subjective domains?

9.5 Normativity, Goals and Negotiation

In what follows, I will discuss Gauker's criticism of Van Deemter, showing that sometimes we cannot avoid referring to what Gauker thinks is a fortuitous coincidence of subjective perspectives. I will stress, on the contrary, that the coincidence of perspectives is not fortuitous; rather, it is the result of a normatively organized work of convergence. This kind of work follows objective rules (norms) given by making explicit the different justifications interlocutors have for their - sometimes diverging - semantic evaluations. I will give three main steps in my argument:

 (i) we cannot always give a clear definition to the goal of a conversation and to the elements of a situation that are normatively relevant;
 (ii) the argument for the necessity of a normative stance needs a distinction between external and internal norms which is not completely satisfying;
(iii) if we do not have a normative context, we may find normative clues to follow in order to disambiguate misunderstandings.

(i) *Normativity and goals*

Certainly a dialogue typically has a theme or a focus around which the dialogue and its domain of discourse develops. Stalnaker speaks of "subject matter" and Gauker of "objective context", which I have translated as "normative context". As we have seen in § 9.3, Gauker defines normative context as what is coherent with the goals of the conversation. Normative context is the set of information relative to the goals of the conversation that speakers should take in choosing their words and hearers should acknowledge in listening to them. In making examples, Gauker refers to specific goals around which to organize such a set of presuppositions, which constitutes the normative context, goals such as - for instance - getting clean water for cooking, and so on. This idea has a certain degree of intuitiveness, and Grice in his "Logic and Conversation" has also used it[11]. An analogous attempt is normally used in other domains of research, like problem-solving or contextual reasoning in artificial intelligence. A good approximation of a normative context is what is sometimes called "working context", that is, the set of information which is imported in a cognitive space to solve a problem, and which can change depending on new facts. For instance, to solve the problem of a journey from A to B, we import in the working context only the basic information about what is necessary to organize the journey (that is, information about acquiring the tickets, scheduling of flights, and not much more). But if something relevant happens (e.g. a ticket is lost) we have to import new information needed for solving the specific problem (see for instance Bouquet and Giunchiglia (1995)). Obviously there are differences between the idea of working context and the idea of normative context. While the working context is the limited set of information needed to solve a problem, the normative context is the set of assumptions needed to understand a dialogue. Both ideas however are linked to individuating some normative means to give boundaries to the set of presuppositions of a discourse or problem-solving situation. They try to give a characterization of what is "needed" with respect to assumed and explicit goals.

The ideas of working context and of normative context are therefore highly welcome in institutional situations, where goals are well-defined. However, a generalization of the idea of normative context to all situations of dialogue runs the risk of giving a misleading picture of what is really going on in communication. Actually, in normal dialogue situations, it is not always easy to decide what the goal of the conversation

---

[11]However Grice himself notes that sometimes conversation does not have a precise goal.

is. What is the fundamental goal of the dialogue between Tommy and Suzy, which might help us to decide the normative context? Reading the example of Tommy and Suzy, one may think that the example does not properly fit Gauker's theory of goal-relative objective context. Gauker's original idea of context (1998, 1999) is a set of propositions that the interlocutors ought to *take* in choosing their words, and ought to acknowledge relative to the goals of the dialogue. Now, we may ask, what Tommy and Suzy *ought* to take the propositions to be? Which is the goal *of the conversation*? Difficult to say, unless we restrict the goal to mutual recognition of intentions and presuppositions. But this does not seem to be Gauker's point, who thinks of specific goals; in the case of Tommy and Suzy he defines "keeping the peace" as the pertinent goal of the conversation (Gauker (1998a), p. 447). But this is quite arbitrary; it is not always easy to decide what the goals of a conversation are, and this example does not help. Maybe both Tommy and Suzy wanted to fight instead of keeping the peace; or maybe they had as a goal showing off their superiority (in terms of marble ownership). Or they just wanted to find an excuse to quarrel, or to show their preference for red marbles over white marbles, without giving any importance to the actual ownership. Who can *decide* what the relevant goals of a conversation are?

What is the case when we face a conversation without specific goals, or with an indefinite set of goals, or different goals for each participant? Or, think of small talk: what is the goal of small talk, besides just talking? The example of Tommy and Suzy may be a bad example[12] for clarifying the point of the idea of normative context, which - as I said - is a good idea in more defined situations. I find difficult to accept a *generalization* of the idea of normative context for treating discourse context and communication. To make the point more explicit, I will examine about two other examples given by Gauker.

In the first example, I say to Alice "I will meet you in front of the department store at noon". After a while I am waiting at Nordstrom's and she is standing in front of Bloomingdale's. Apparently we had two different presuppositions (and two different interpretations of my words). It is a clear case of the presupposition coordination problem.

---

[12] We have a short piece of dialogue: what can we do with this piece of text alone? We lack much relevant information possessed by the speakers and we must guess. A piece of text is always cut off from what precedes and what follows. We may speak of Tommy's initial statement as not so correctly assertable in the context only if we abstract from previous linguistic exchanges of the two speakers. We may always consider a possible larger context to make more rationale the assertions of the speakers and their goals.

To avoid recourse to infinite reflexive thinking[13] ("I will go where she thinks I think she will go", and so on), Gauker suggests referring to the objective-normative context: we need to refer to an "objective" context, which is decided by some relevant objective features of the situation. In this way, Gauker explains the lack of coordination referring to a disparity between what the objective-normative context really did contain and the take on the context of one of the two. Simply, one of the two has a *mistaken* presupposition. But what is the objective-normative context here, what should decide the right and wrong presuppositions? My sentence may be honestly capable of different interpretations, and it is easy to imagine a situation in which there is no way to decide which is the "real" objective or normative discourse context. It is possible to imagine cases in which both Alice and I may have *good reasons* to believe that respectively her or my interpretation is the correct one. In this case, it appears difficult to decide which is the "real" or "objective" or "normative" context, except by referring to some inner intentions (probably of the person who has a stronger social status). Arló-Costa (2006) develops this point further, and I will not pursue it here.

Gauker's second example is different and more refined: referring to C getting in a rather beat-up car, A says to B "Her car was stolen". What is the intended presupposition? Two possibilities are at stake:

(i) C is getting in her car, which is now damaged after having been stolen.

(ii) C is getting in a used car, because her own car has been stolen and lost.

Let us assume that (ii) holds. How can B decide which of the two possibilities is the right one? Simple enough, if the situation is (ii) B *should* presuppose (ii). Given the general theory, the contents of the normative context cannot by definition be confined to what is only "contingently" subsumed under the speaker's assumptions. The speaker too may be mistaken. A might have forgotten which car C was using and erroneously he may believe (i). So B, if he has reasons to think that A presupposes (i), also ought to recognize that A does not presuppose what belongs to the normative context. Therefore B has, in this peculiar case, two "oughts" and not only one: he should presuppose (ii) and he should recognize that A is mistaken in his presupposing (i). Poor B! How heavy normative responsibility is in order to explain the lack of

---

[13]On the presuppositrion coordination problem see also Arló-Costa (2006). There are psychological restrictions about the number of reflections possible for humans. The possibility of an infinte regress is a logical possibility, not a psychological one. We should therefore be very cautious on giving too much stress on this criticism.

presupposition coordination with A!

Here it seems to me that Gauker puts the cart before the horse. Explaining the misunderstandings through the reference to the real situation is an unnecessary step. We may do much work on understanding communication and its misunderstandings even *before* considering either objective reality or what objective reality might be. In fact, in order to understand what is going on in the dialogue, we are not interested in what is really going on with the ownership of the car. On the contrary, we need to understand what is going on in the relative presuppositions of the two interlocutors, even without knowing the real state of affairs. Actual reality here is a matter independent of the dialogue, and needs further evidence (which can be provided by some other step in the conversation, maybe asking the relevant girl, who is the most reliable subject in this case, to tell which assumption is the right one). The information that (i) or (ii) is the real case, is irrelevant for assessing the relative misunderstandings of the two participants in the dialogue, in the case that A believed (i) and B believed (ii). When we try to understand a dialogue of this kind, we are not necessarily interested in the correctness of what the speakers believe, but in the source of their misunderstanding. We may push the case further by imagining a situation in which it is not possible to decide which is the real state of affairs. Simply imagine that C, the girl, abandons the scene and never comes back. Still, we may understand the misunderstanding which is going on in the dialogue and find a means by which to extract the different presuppositions of the two speakers (for instance, they make different inferences).

Certainly B should look for what the case is; he should not look just for what A thinks the case is; but misunderstandings among the two concern their beliefs, not the real state of affairs. We may easily assert that B *should* presuppose what is really the case. However this assertion is empty unless it is possible to have access to the real situation. If it isn't, the wheel is idling.

(ii) *External and internal norms*

These short remarks on the examples bring us to discuss where and how Gauker makes the trick that permits him to put the cart before the horse. I think that we need to analyze the first step in his argument for normative context, that is the distinction between external and internal norms given in Gauker (1998b). *External* norms tell us what a speaker should think, given that the world works in a certain way; *internal* norms tells us what a speaker should think, given that he or she has certain beliefs. If the norms are external, discourse context is given by

shared assumptions; if norms are internal, discourse context is given by what the interlocutors suppose to be shared assumptions. But internal norms are "merely subjective reflections of external norms"; therefore, in treating context dialogue, we should take external norms into account. Gauker puts forward an example: it would be inappropriate to say "Matt knows that his paper is late" unless the interlocutors share the assumption that Matt's paper is late. The example should give support to the idea of an external norm, and something in it is certainly convincing. Given that "to know" is a factive verb, which is the most typical presupposition trigger, it is inappropriate to say "Matt knows that his paper is late" unless the speaker presupposes that (i) Matt's paper is late and (ii) his hearer assumes it too. Until now it is very difficult not to agree. But there is a further step I would not take: the step is to claim also that, for an assertion to be appropriate, we need to require that speaker and hearer *share* the same presupposition. It seems to me that in making this delicate step we begin to put the cart before the horse. I will try to explain my worries related the defeasible status of our assertions.

Our assertions are made with some ground or justification, but they are always defeasible. It is appropriate to say that Matt knows that his paper is late if there is enough evidence that his paper is late. The speaker might know the deadline, therefore he infers that Matt's paper is late, and - having asked Matt about the matter - has received the answer that he knows that. Shall this scenario be enough to ascertain the appropriateness of the speaker's speech act? I suppose we cannot avoid it. However it is easy to imagine a further scenario, where a hearer knows that the deadline has been delayed, and therefore she does not share the presupposition that Matt's paper is late. She understands the sentence uttered by the speaker and understands his presuppositions. However she has the right to correct the sentence, which is appropriate under certain conditions, but not appropriate in the new situation of the change of deadline. We, observers, cannot know in advance how a situation may change and cannot ask the appropriateness of an assertion to rely on the "real situation", given the intrinsic limitation of our epistemic access to the world. We may still require that what we *mean* by saying "x knows that p" depends also on p being true. There is a difference in saying that x knows that p only if p is true, and saying that my assertion of "x knows that p" is *appropriate* only if p is true. We cannot ask that the *appropriateness* of our assertion about x's knowing something depends on the objective state of affairs, beyond our means of recognizing it. The appropriateness (not the truth) of an assertion is always relative to a certain set of justification conditions we normally

take for granted. But we may find *always* cases where these conditions are no longer fulfilled. Given this general situation we have an alternative: either reject *any* assertion as appropriate because always prone to fallibility, or to accept all our assertions as appropriate under certain conditions of justification. In this case, to judge the appropriateness of an assertion in a dialogue we need to know the set of presuppositions held true by the speaker. These are the presuppositions that the speaker *thinks, with good reason*, hold true also for the hearer. However these presuppositions do not necessarily have to be shared by the hearers for an utterance of the speaker to be appropriate. The hearers are supposed to share the presupposition, but they may not, in case they have further unexpected information. To impose that an utterance is appropriate only in case its presuppositions are shared by speaker *and* whatever hearers is to make a too strong demand, as if the appropriateness of an utterance depended on its objective truth, regardless of time and accessibility conditions. If this were the case, probably no utterance might ever be considered appropriate, given that there might always be, in principle, some interpretation - given by further information or by new cognitive settings - which falsifies the utterance.

Correctly for Gauker "it is vacuous to say that speakers have responsibility to make sure the assumptions that they suppose to be shared really are assumptions they suppose to be shared". But it is not vacuous at all to say that speakers have responsibility to *give reasons* for the assumptions they suppose to be shared, reasons which explain why the assumptions are the right ones, and why the other interlocutors *should* accept them. But, once given that, they are responsible to be ready to recognize that the grounds for their assumptions are mistaken when offered evidence. I mean with this conclusion to stress that the normative aspect does not depend on a previous assumption of an objective state of affairs, but on the honest and rational search for this objectivity.

Gauker sees no way to define internal norms which are not "mere subjective reflections of external norms". I will answer that internal norms are an expression of a fundamental aspect of the limits of human reason or are an expression of the intrinsic defeasibility of our descriptions. Internal norms deal with the defeasibility of cognitive contexts regarding the objective state of affairs. To attain the "objective" state of affairs, we need a description of it. But every description is an expression of a defeasible point of view. We have to enclose this defeasibility into our theory and leave conversational context to be the interaction of points of view, through which we might eventually converge toward what we consider the right one, given its grounds, its reliability and its

justifications.

Normativity is a conditional constraint: "*if* the real situation is so and so, *then* the speaker should say so and so." But there is no warrant for thinking that we always have access to the real situation, or we may find unexpected doubts about a previous description of it. Therefore, even if we do not deny the existence of external norms, we cannot take them to be the foundation of discourse context; contrary to the standard view, they are dependent on internal norms when proper attention is given to the defeasibility of our access to what is the case. This does not mean that what is the case depends on what people think the case is. It means that our *descriptions* of what the case is rely always on a background of practices and beliefs. Therefore this internalist attitude does not mean that what is right is reducible to what is considered right, but just that what is right is always postulated or decided on the background of the open discussion among different points of view about what is right.

(iii) *Normativity and anaphora*

How to deal with a discourse with no definite normative setting, with no clearly definite goal? Let us take again the dialogue between Tommy and Suzy, where eventually Tommy realizes that Suzy is giving an interpretation different from his. How is this convergence realized? What is relevant here appears to be the negotiation about the use of anaphora and demonstratives, given a basic agreement on the use of pure indexicals and proper names. Let us check the final part of the dialogue:

(1) T. Yes, *they* are [mine]. Mom gave them to me.
(2) S. Dad gave me *those* marbles and Mom doesn't even know I have them.
(3) T. I'm not talking about *those* marbles; I'm talking about the marbles in *my* room.

The anaphoric chain, which has been used by Tommy and Suzy as if it referred to the same objects, breaks at the third sentence in the above piece of dialogue, when Tommy reacts to Suzy's claiming that "Dad gave me those marbles." Realizing the different connection of marbles with Dad and Mom, Tommy makes it explicit that the anaphoric use of "they", "them" and "those" is ambiguous. He then begins to bargain over the interpretation of the anaphora and *accepts Suzy's use* of "those" in order to make *his* point. The problem is: how and when did he detect the different use of "those"? On the one hand, Suzy probably accompanied "those" with a gesture, linking "those" to a pure indexi-

cal "here" (unarticulated constituent?). On the other hand, Suzy made explicit reference to a different (causal) origin of the intended reference of the term "those": Dad and not Mom. Recognizing the contrasting use of "those" made by Suzy, Tommy makes explicit the different domain of quantification of the relative uses of "they/those" and clarifies that his use was not directed to the domain of quantification chosen by Suzy. In order to do that, Tommy has to take a step backwards and become an external observer of the dialogue. Making this step he begins to bargain the domain of "those" and "they".

We have given the case in which he accepts Suzy's use of "those". However he might also have pretended that his use was the only correct one and that Suzy is mistaken. In this case his payoff in the bargaining would be a psychological superiority over Suzy, and the outcome would be that his interpretation, or the domain or the proposition, would be unique, the one *he* intended. In our example, on the contrary, he chooses the most efficient way, which reaches an efficient result with the least effort: he accepts Suzy's interpretation showing that his intended referent of the domain of quantification is different from Suzy's. He reaches a fair solution (we refer to two different domains) from incompatible premises (what I say is true and what you say is false).

In this case semantic bargaining[14] employs basic features like demonstratives ("*those* marbles"), and works on the control about the interpretation of anaphoric chains ("they"..."those"). To disambiguate demonstratives and anaphoric chains the speakers use proper names ("Dad", "Mom") and pure indexicals ("I", "me", and an implicit "here"). Through the concordance and discordance on the use of these devices, the interlocutors realize that their two anaphoric chains, where they both use the term "they", have different anaphoric intended initiators and need to be recognized as two different chains. In such a way they realize that there are two interpretations of the "red ones", neither of which is the "right" one.

In using and comparing the different uses of demonstratives and indexicals they are *compelled* to recognize the existence of different domains of quantification. As I have suggested in (i), if a dialogue has no precise goal then there is no particular proposition or set of propositions people *should* share, besides certain agreement on personal identity and the use of indexicals. We cannot say that Tommy and Suzy should share the proposition given by Tommy's interpretation, because the proposition given by Suzy's interpretation is perfectly coherent with

---

[14]This analysis and the terminology used has been influenced by Massimo Warglien, who analyzes with more detail the requirements necessary to speak properly of "meaning negotiation".

the actual situation. But there is also no reason why Tommy should share Suzy's interpretation, having begun the dialogue with his own interpretation. Therefore the problem is not *which proposition* should be shared, but *which means* the interlocutors have to recognize the different propositions at stake.

Let us suppose that Suzy said: "*It* is not true" - as her first reaction to Tommy's utterance ("all the red ones are mine"). To this Tommy could easily reply: "No, it is certainly true.". What does "it" refer to in the two cases? A normal anaphoric analysis requires that there is a *unique* referent for "it", which - in an expressivist view - should be a proposition. *Which* proposition? If *we* have to describe it, we should make the proper embedding, making it explicit that Suzy refers to what she thinks Tommy has in mind; therefore "it" is ambiguus between "what Susy believes Tommy believes" and "what Tommy believes". But we know that Suzy believes that Tommy believes that all Suzy's red marbles are Tommy's, or, using our old way of expressing:

$$\text{Suzy believes that Tommy believes that}$$
$$\forall\, x \in \mathrm{DoD}_\mathbf{s}\ (\mathrm{RedMarble}\,(x) \longrightarrow \mathrm{Tommy's}\,(x))$$

In this case the ambiguity is given by the anaphoric use of "it", where two tokens of the same type (the pronoun "it") are used by Suzy and Tommy to refer to two different propositions, or different interpretations of the same utterance. Here the "it" is like a name "Cicero" used ambiguously both for the Roman orator and for another historical character, like the case in which two different persons - in front of an utterance of "he is coming", interpret "he" as referring to two different individuals in the scene. We need to clarify the strategy people use to point out where the use of two different anaphoric chains starts from, as basic tool for disambiguation. The relevant point is therefore to analyze the basic steps needed to disambiguate misunderstandings when propositions are the referents of indexicals and pronouns.

9.6   Conclusion

When no definite goal is at stake, we have no means by which to decide which is the intended interpretation (the domain of discourse), but we have to study the structure of semantic negotiation, where speakers are forced to understand different relevant interpretations. Checking the concordance and discordance in the use of anaphoric chains is a basic means which compels us to realize the possibly different interpretations, or different local domains[15]. Speakers do not need to converge

---

[15]I am suggesting that underspecified logical languages and local model semantics (e.g. Giunchiglia-Ghidini 1999) might be expressively equivalent.

towards *one* interpretation, but recognize the different interpretations as just different, and not necessarily competing. The rules which govern the admitted substitutions in an anaphoric chain are ways to compel people to find agreement and disagreement. These rules compel people to make explicit the different commitments they have on the content of their assertions, which may impinge upon different tokens of the same type. When two tokens of the same type are referring to different domains, misunderstanding is easy to discover after some step in the conversation. However, the normativity that permits communication to be successful cannot always be given in advance, in a theory of what should be shared. Other norms may help, and they are laid down in the rules which compel the interlocutors to recognize the different commitments (e.g. different inferences) on the expressions used.

This does *not* mean that we must accept an expressivist picture of communication as devoted uniquely to transmitting something mental from one mind to another mind. Considerable effort is devoted to inferential work, checking the differences in the use of tokens of the same type, looking for the point where the divergence of anaphoric chain becomes explicit. The inferential work arrives, eventually, at the construction - step by step - of a linguistic setting which compels interlocutors to get at the recognition of different interpretations given by the different theories at stake.

Maybe that is why Gauker does not give so much attention to the supposed difficulty of the expressivist facing the neutral position where *both* interpretations are right. Actually, this is the easiest result of the dialogue, where - as a result of a negotiation - Tommy and Suzy recognize the point of view of the other, saving in this way the truth of the apparent contrasting assertions: "?It is true that? they are mine" vs. "No. It is not true". They are both true, but two tokens of the term "it" have different referents. The solution of the puzzle is very simple[16], but the strategy to arrive at the solution impinges upon a very sophisticated level of linguistic and contractual abilities, which compel speakers to recognize that their own points of view are not the unique

---

[16]The example under consideration is a typical case of confusion about identity. Here two tokens of the same expression ("they") have two different referents (two different sets of marbles). The case might be more complex if we build up a case where the intended domain is the same, but the epistemic access to it is different. In this case, the disagreement would seem (erroneously) a disagreement about the facts, because the assertions are interpreted on the same class of marbes. To solve the problem, we might speak of difference of *intensions*; but also we may say that the same class of red marbles is considered under two different points of view, which can be expressed in two *theories* or local contexts. The negotiation needs just a further step.

ones on the market.

References

Akman, Varol. 2006. *Perspectives on Context*. CSLI.

Arló-Costa, Horacio. 2006. *Perspectives on Context*. CSLI.

Bach and Kent. 1996. The semantics-pragmatics distinction: What it is and why it matters. In K. Turner, ed., *The Semantics-Pragmatics Interface from Different Points of View*, pages 65–84. Oxford, Elsevier.

Benerecetti, M., P. Bouquet, and C. Ghidini. 2006. *Perspectives on Context*. CSLI.

Benerecetti, M., P. Bouquet, and S. Zanobini. 2003. Verso una logica cognitiva per i linguaggi indicali. *Epistemologia* 26(2):201–236.

Bianchi, C. 1999. Three forms of contextual dependence.

Bouquet, P. and F. Giunchiglia. 1995. Reasoning about Theory Adequacy. A New Solution to the Qualification Problem. *Fundamenta Informaticae* 23(2–4):247–262.

Brandom, R. 1994. *Making it Explicit*. Duckworth, London.

Davidson, D. 1984. *Inquiries into Truth and Interpretation*. Clarendon Press, Oxford.

Davidson, D. 1986. A nice derangement of epitaphs. In E. Lepore, ed., *Truth and Interpretation. Perspectives in the Philosophy of D. Davidson*, pages 433–446. Blackwell, Oxford.

Gauker, C. 1994. *Thinking out Loud*. Princeton University Press.

Gauker, C. 1997. "domain of discourse. *Mind* 106:1–32.

Gauker, C. 1998a. Intelligibility in semantics: Reply to van deemter. *Mind* 107(426):447–45.

Gauker, C. 1998b. What is a context of utterance? *Philosophical Studies* 91:149–172.

Ghidini, C. and F. Giunchiglia. 2001. Local models semantics, or contextual reasoning = locality + compatibility. *Artificial Intelligence* 127(2):221–259.

Giunchiglia, F. 1993. Contextual reasoning. *Epistemologia, special issue on I Linguaggi e le Macchine* XVI:345–364.

Hinzen, Wolfram. 1999. Contextual dependence and the epistemic foundations of dynamic semantics. In P. Bouquet, L. Serafini, P. Brézillon, M. Benerecetti, and F. Castellani, eds., *Modeling and Using Contexts: Proceedings of the Second International and Interdisciplinary Conference, CONTEXT'99*, LCNAI 1688, pages 186–199. Berlin: Springer-Verlag.

Kamp, H. 1993. *From Discourse to Logic*. Kluwer, Dordrecht.

Kaplan, David. 1978. On the logic of demonstratives. *Journal of Philosophical Logic* 8. reprint in Themes from Kaplan (see below).

Kaplan, D. 1989. Afterthoughts. In J. Almog, J. Perry, and H. Wettstein, eds., *Themes from Kaplan*. Oxford U.P., Oxford.

Lewis, D. 1969. *Convention: a philosophical study*. Cambridge (Mass).

Lewis, D. 1980. Index, context, and content. In S.Krange and S.Ohman, eds., *Philosophy and Grammar*, pages 79–100. Reidel.

McCarthy, J. 1993. Notes on Formalizing Context. In *Proc. of the 13th International Joint Conference on Artificial Intelligence*, pages 555–560. Chambery, France.

Penco, Carlo. 1999. Objective and cognitive context. In P. Bouquet, L. Serafini, P. Brézillon, M. Benerecetti, and F. Castellani, eds., *Modeling and Using Contexts: Proceedings of the Second International and Interdisciplinary Conference, CONTEXT'99*, LCNAI 1688, pages 270–283. Berlin: Springer-Verlag.

Penco, C. 2000. Three alternatives on context. In D. Marconi, ed., *Knowledge and Meaning. Topics in Analytic Philosophy*. Mercurio, Vercelli.

Perry, J. 1998. Indexicals, contexts and unarticulated constituents. In *Proceedings of the 1995 CSLI-Amsterdam Logic, Language and Computation Conference*, pages 1–6. CSLI Publication, Stanford.

Recanati, F. 2001. On what is said. *Synthese* 128:75–91.

Recanati, Franois. 2002. Unarticulated constituents. *Linguistics and Philosophy* .

Recanati, F. this volume.

Sbisà, M. 1999. Significato, interazione, testo. In P. Basso and L. Corradin, eds., *Eloquio del senso*, pages 260–276.

Searle, J. 1983. *Intentionality*. Cambridge U.P., Cambridge.

Stalnaker, R. 1999. *Context and content : essays on intentionality in speech and thought*. Oxford Unviersity Press.

Travis. 1997. Pragmatics. In B. Hale and C. Wright, eds., *A Companion to the Philosophy of Language*. Blackwell.

Van-Deemter, K. 1998. Domains of discourse and the semantics of ambiguous utterances: A reply to gauker. *Mind* 107:443–445.

Van-Deemter, K. and S. Peters. 1996. *Semantic Ambiguity and Underspecification*. CSLI pulications, Stanford.

Wittgenstein, L. 1953. *Philosophical Investigations*. Blackwell, Oxford.

10

Fictional Contexts

Andrea Bonomi

10.1 The problem

In event semantics the content of sentences such as:

(1)   James Joyce shaved on the platform of the Martello tower

and

(2)   Andrea Bonomi wrote a letter to the dean on April 4, 1984

is accounted for, among other things, in terms of particular relations between events (or states[1]) and places or times. Roughly speaking, an event α is said to *occur* in a place p (or interval t) if the spatial (temporal) extension of α is located in p (or t). Let the predicate 'Occ' denote such a relation. From this point of view, part of the content of the above sentences can be associated, respectively, with formulas such as:

(1')   Occ(f, the platform of the Martello tower)
(2')   Occ(e, 04.04.84)

where f is the event of Joyce's shaving and e is the event of Andrea Bonomi's writing a letter to the dean.

I presented elsewhere an argument[2] to the effect that the content of

---

[1]In the present paper I will concentrate on events proper and I will ignore states. But this point has no relevance with respect to the general issue I am going to discuss.

[2]This argument, based on the "downward indeterminacy" of fictional entities and events, is discussed in Bonomi and Zucchi (2001).

true[3] sentences like:

(3)   Buck Mulligan shaved on the platform of the Martello tower
(4)   Winston Smith started writing his diary on April 4, 1984

cannot be reconstructed in terms of a simple relation of occurrence between a fictional event such as Mulligan's shaving or Smith's writing his diary and a particular place (the platform of the Martello tower) or a particular time (April 4, 1984).

In general, saying that an event α occurs in some place p or in some interval of time t entails some systematic relations between *parts* of α and *parts* of p (or t). For example, if the event g of Joyce's propping a mirror on the parapet of the platform is a proper part of the event f of Joyce's shaving, and if u is the spatial extension of f, then there must be a proper part v of u such that v is the spatial extension of g. In the same way, if the event d of my writing the address is a proper part of the event e of my writing a letter to the dean, and if u is the temporal extension of e, then there must be a proper part v of u such that v is the temporal extension of d. Using *Ext* to denote a function which assigns spatial (or temporal) extensions to objects or events and '$\subset$' to denote the relation *part-of*, it is possible to state the following general principle:

(PE) $Ext(\alpha) = u$ & $\beta \subset \alpha \rightarrow \exists x[Ext(\beta) = x$ & $x \subset u]$

In the case of events, (PE) means that an event β which is a proper part of an event α must have a spatial (or temporal) extension which is a proper part of the spatial (or temporal) extension of α.

Intuitively speaking, principle (PE) expresses a necessary condition for an event α to have a spatial (or temporal) extension u: a condition motivated by the idea that the extension of α is determined by its constitutive parts, i.e. by the events that compose α. So, what the "downward indeterminacy" argument is intended to show is that whilst real events always satisfy this condition, fictional events, as a rule[4], are unable to satisfy it.

Consider, for instance, examples (3) and (4). These are privileged situations, because there is an explicit reference to a certain object

---

[3]In view of what is narrated, respectively, in *Ulysses* and *1984*.

[4]This qualification is important because the kind of argument I am going to discuss is not intended to prove that it is *always* impossible to attribute a definite extension (in the spatio-temporal context of real events and objects) to fictional events, but that this is the default case at least when we have to do with novels or related cultural artifacts. As shown here, this is so even in the rare situations where the text refers to a place or time which is exactly specified. A peculiar problem is raised by fictional events described in a play or in a film.

(which is a real tower) or a certain date (which is a particular day), so that one might say that the spatial (or temporal) extension u of this object (or day) is the spatial or temporal extension of the fictional event α at issue. The problem is that, as we will see in a moment, it is not possible to attribute, in the world which surrounds us, an appropriate extension to the subparts of α. Therefore, because of such a violation of principle (PE), no real extension, in this world, can be attributed to α and, strictly speaking, one cannot say that α occurs in u, i.e. the spatial (or temporal) extension, in our world, of the relevant place (or date).

For example, suppose that in Orwell's novel it is specified that Smith writes the first three pages of his diary on April 4, 1984, exactly between 3 and 7 p.m. So, at least in this case, it would seem that there is no difference between the (fictional) event of Smith's writing the first three pages of his diary and the (real) event of Andrea Bonomi's writing a letter to the dean: the temporal extensions of both events might be conceived of as particular intervals on the time axis.

The problem is that in the case of Smith's diary it is *in principle* impossible to determine the temporal extension of the constitutive parts of the event. Take for instance the event of Smith's writing the first page of his diary, which is part of the global event of Smith's writing the first three pages. What is its temporal extension? No answer is possible, and this is so not because of our ignorance, but for the simple reason that no specification is given by the text. This "downward indeterminacy" of temporal qualifications, *which is just an aspect of a more general phenomenon of indeterminacy*, is an essential feature of fictional events or entities: by making our analysis of the internal constituency of these events or entities more and more specific, as a rule we reach a level at which no temporal extension can be attributed to some subpart of the event or entity at issue, and this is a violation of principle (PE).

Since a similar downward indeterminacy affects the spatial characterization of a fictional event or entity x, we can conclude that if a particular time or place in the world around us is selected as the temporal, or spatial, extension of x, a paradox arises: this event or entity would turn out to have a temporal, or spatial, extension even though its constitutive parts have no extension. Once more, this is a violation of principle (PE).

A consequence of this argument is not, of course, that the event e of Smith's writing the first three pages of his diary has no temporal extension *in* the world of Orwell's novel or that the event f of Buck Mulligan's shaving has no spatial extension in the world of Joyce's novel, but only that, unlike the real events described in (1) and (2), e

and f do not have such extensions in our world. Anyway, the argument is sufficient to show that, whilst in the case of (1) it is perfectly appropriate to speak of a genuine relation of occurrence between the event of Joyce's shaving and a real place, such an attribution is problematic in the case of (3). Strictly speaking, it is misleading to refer to a simple relation of occurrence between a fictional event (such as x's writing a diary or y's shaving) and a particular place or time (such as a tower or a given date). Yet, in view of the intuitive truth of a sentence like (3) or like *Sherlock Holmes lives in London*, there must be some relation between the event (e.g. Mulligan's shaving or Sherlock Holmes' living in London) and a real place (e.g. the Martello tower or London). The idea is that this relation is mediated by the existence of the story in question. This is why I will speak, in these cases, of a relation of *scene-setting* between an event, a story and a particular place or time. In what follows I will try and characterize this relation.

10.2   Types of assertion

A crucial role is assigned, in my reconstruction, to the principle of importation. The intuitive idea is that the use, in fiction, of familiar designators is based on the background information associated with them: the implicit assimilation of this information helps to set up a suitable frame for the story. This is what happens, for instance, in the case of designators referring to places (as the proper name *London*), to times (as the date *April 4*), to events (as the definite description *The Borodino Battle*), to persons (as the proper name *Napoleon*), and so on.

Before addressing this problem, let us reflect for a while on a different class of designators which occur in fictional stories. Consider, for example, Proust's *Recherche*, where we find proper names such as *Combray* or definite descriptions like *Le Grand-Hôtel de Balbec*. On the one hand no real place is the bearer of these names (although, as we will see, this is no longer true of the name *Combray*), whilst, on the other hand, it is easy to identify which real places can be associated to these names. These places are, respectively, Illiers and Le Grand-Hôtel de Cabourg. In general, this kind of identification is made possible by resorting to the background information concerning these real places (including the role they played in the creation of the story), which allows us to establish systematic relations between fictional entities and real referents.

More precisely, given a story H and a set of properties X, selected among those which are assumed to characterize an individual α, I will

speak of a function g which, thanks to the properties in X, associates a character[5] β to α. Thus, a statement of the form:

(5) $g(\alpha, H, X) \;=\; \beta$

means that β is the character which, in the light of the story H and the relevant properties in X, corresponds to the individual α. (In what follows I will often speak of g as a one-place function, under the assumption that the reference to H and X is implicitly fixed by the context).

To make this point clearer, consider the distinction, made in Bonomi (1987), between three different types of sentences having to do with fiction. First of all there are "textual" sentences, i.e. sentences which are part of the text itself, as for instance:

(6)   M. Vinteuil s'était retiré auprès de Combray.

We will see how these sentences are treated in Frege's theoretical framework. For the moment, I shall confine myself to observing that when we run into a textual sentence like (6) in the process of reading the *Recherche* we hardly ask ourselves whether Proust (or, more precisely, the narrator) is saying anything true or false. On the other hand, a fictional story, in so far as it is transmitted by a text within a community, can be considered as a particular context of information which allows *us* to make true (or false) statements with respect to the content of the story itself. This is what happens, for example, when *we* say to a friend of ours:

(7)   Vinteuil lives in Combray.

By means of such sentences (that I will call *paratextual* sentences) we can state something true (or false) on the basis of the story narrated in the *Recherche*. In general, the idea is that paratextual sentences refer to a given context of information, provided by the story. Once this contextual reference is taken into account, we obtain the intended interpretation of a sentence such (7), that is:

(7')   (In the *Recherche*) Vinteuil lives in Combray.

To account for this interpretation in a suitable theoretical framework, let us adopt a formal language like the one illustrated in Bonomi (1977, 1979), where particular indices[6] are used to refer to contexts.

---

[5]This term is used in a broad sense in order to include fictional places or times, rivers, animals, etc.

[6]The role of indices, as variables over contexts, is discussed in the appendix. A different treatment of the prefix 'In the story X,' based on the analysis developed in Lewis (1978), is presented in Ross (1997). The semantics of this prefix is also

So, if A is a sentence and D an index referring to a story (the story narrated in the *Recherche*, in our example), the sentence '$(A)_D$' can be roughly paraphrased as 'In the context of the story D we have that A'. (It should be noticed that this is not an *ad hoc* move: this kind of reference is a general pragmatic phenomenon, as shown in Bonomi (1998).) For example, under the interpretation illustrated by (7'), in such a formal language (7) would be associated to this formula:

(7") $[\text{lives-in(Combray, Vinteuil)}]_R$

where 'R' is an index which refers to the *Recherche*. (7") can be read as follows: In the context of the *Recherche* we have that Vinteuil lives in Combray.

The problem, with this kind of approach, is that it is essentially based on the use of an index to fix the implicit or explicit reference to the text or to the story associated with that text: so, it can be applied *only* to paratextual sentences, i.e. sentences that we might utter to make a report about that story or text. Yet, this solution is not available in the case of genuine textual sentences such as (6), which of course do *not* occur in *our* talks about texts or stories. Moreover, there is another kind of sentence having to do with fictional entities which cannot be accounted for by the use of indexed sentences. This is the case of *metatextual* sentences like:

(8)   Orson Welles loves Don Quixote

In fact, what we state here is simply something true *in the context of the real world* (given Orson's well-known passion for this character), not something true in the story. In this case, the text and the characters that it generates do not represent the relevant context of information, but they are mere *objects* of discourse (exactly as other cultural artifacts[7], like a symphony or a statue). As a matter of fact, the truth of (8) does not depend on the information provided by the story: we can utter (8) truthfully even if we know nothing about the content of the novel. As a consequence such sentences cannot be interpreted in terms of paratextual sentences like (7): once more, indexing (as used in the case of (7)) is not a solution. In principle, metatextual sentences and paratextual sentences must be distinguished, for mixing these two different levels might give rise to very odd statements. For instance, let

---

discussed in Currie (1990).

[7]In Kripke (1973), characters are presented as entities existing in the context of our world (in virtue of the existence of the relevant stories). An analysis of characters as cultural artifacts is sketched in Bonomi (1994). For a full analysis of this notion see Thomasson (1999).

us suppose that someone asks us:

(9)   Who loves Don Quixote?

This question, in its *paratextual* interpretation, might be followed by an answer such as:

(10)   Sancho Panza loves Don Quixote

In its *metatextual* interpretation it would be associated quite naturally to an answer like (8). Yet, the answer:

(11)   Orson Welles and Sancho Panza love Don Quixote

would sound very odd because two different levels of discourse are confused. The intuition is, of course, that Sancho Panza loves, *in the story*, Don Quixote as a person, whilst what Welles loves, *in the real world*, is a character, i. e. an entity generated by the text. In other terms, Welles can love a character exactly as he can love a brand of cigars. Both the character and the brand of cigars exist in the real world (unlike the person Don Quixote, who does not exist in this world): the latter as a result of a material production, the former as a cultural artifact generated by a linguistic activity culminating in a text.

10.3   Complex types of assertion

Unfortunately, things are not that simple, for there are sentences about fiction which call for some refinements of the above analysis. To see the complexity of the problem, consider the following sentence:

(12)   In the *Recherche* Miss Vinteuil's father is a very shy piano teacher and an underestimated composer. Proust created this character after studying several musicians.

The definite description *Miss Vinteuil's father*, in the first sentence, is the antecedent of the anaphoric expression *this character* in the second sentence. But notice that the definite description in the first sentence occurs in a paratextual context, because it is used to mention certain properties that a *person* has in the light of what is narrated in a fictional story, whilst in the second sentence the anaphoric expression occurs in a metatextual context, because it is used to mention certain properties that a *character* has in the light of historical information. Indeed, we get an equivalent statement if the anaphoric expression *this character* is replaced by the antecedent expression itself:

(12a)   In the *Recherche* Miss Vinteuil's father is a very shy piano teacher and an underestimated composer. Proust created Miss Vinteuil's father after studying several musicians.

This means that proper names and definite descriptions can be used

to speak of persons (or rivers, animals, etc.) in the context of a report about a fictional story or to speak of characters, as cultural artifacts, in the context of real facts.

A natural explanation of the double role of singular terms in sentences about fiction is based on the idea that they can have *different functions in different contexts*. More exactly, we can use a proper name or a definite description to speak of a person (or animal, river, town, etc.) seen against the background of the information provided by the text (and this is the case of paratextual sentences) or to speak of a character seen against the background of the information concerning real facts (and this is the case of metatextual sentences). For instance, in the first sentence in (12a) the reference to the contextual information **R** associated with the *Recherche* allows us to select as the world of evaluation a counterfactual situation s where there is a *person* who is Miss Vinteuil's father and who is a very shy piano teacher. As for the second sentence in (12a), an implicit context shift is required to determine its content: what is relevant, this time, is the background information **B** concerning the actual world w_0. And in *this* context the definite description *Miss Vinteuil's father* (or the proper name *Vinteuil*) is used to speak of a character *stricto sensu*, not of a person.

To sum up, in both cases the *same* proper name is involved (this is why an anaphoric relation is possible in sentences like (12)), although its two roles in different contexts are distinct, as witnessed by the oddity of (11) (where, as we have seen, this distinction is not respected) and by the ambiguity of sentences such as:

(13)   Vinteuil is underestimated

which can be followed either by this kind of explanation

(13a)   Even his friends, in Combray, do not know that he is a great composer

or by this other argument:

(13b)   Literary critics often ignore that this character is very important to understand Proust's ideas about music.

For similar reasons, the following sentences can be consistent (even though they seem to contradict each other):

(13)   Vinteuil is underestimated

(explanation: this is what we know from the *Recherche*, as stated by (13a))

(14)   Vinteuil is not underestimated

(explanation: (13b) is false)).

In short, we can speak of a character as a *person* (or a town, a river, an animal, etc.), if the relevant context is the body of information provided by the text, or we can speak of a character as a character *stricto sensu*, i.e. as a cultural artifact, if the relevant context is a body of information concerning empirical facts. A single designator occurs in both circumstances (e.g. a proper name such as *Vinteuil* or a definite description such as *Miss Vinteuil's father*), but this noun phrase is associated with two different roles, for it can be used to describe the character "from within" (i. e. in terms of properties that persons, rivers, animals, etc., have in the story), or "from outside" (i.e. in terms of events involving this character in the world around us).

It should be noticed that the first notion of character is not problematic from a semantic point of view, for it involves familiar kinds of individuals, such as persons, animals, rivers, and so on. As we shall see when discussing indexed sentences, what we have to do is simply to point out the *restrictions* which govern reference and quantification in this case. (Roughly speaking, the idea is that only de dicto structures are appropriate here, in order to seal the existence of individuals such as persons, animals, rivers, etc. within a "modality", as suggested in Prior (1968, p. 143), that is within the context of the counterfactual information provided by the story[8]).

What about characters *stricto sensu*, which is the second notion we have just introduced? In this sense, characters do exist in the context of *our* world (as cultural artifacts), and they do *not* coincide with familiar entities like persons, animals, rivers and so on. So, how can they be analyzed from a semantic point of view? Let us consider again sentence (8), repeated here:

(8)   Orson Welles loves Don Quixote

and let us compare it with this other sentence:

(15)   Orson Welles loves his wife, Rita Hayworth.

Whilst it is quite natural to say that in the case of (15) the second relatum of the relation at issue is a person, such a statement would be problematic in the case of (8), because there is no such person. Probably, what Orson Welles admires or likes is a *type* of person, whose characteristics or properties are fixed by the text. And the same can be said of the second part of sentence (12b): what Proust created is not, of course, a person like you and me, but, once more, a *type* of person, with such and such properties described in the *Recherche*. To "create" a character, in this sense, is nothing but to compound properties of

---

[8]I will address this issue in the Appendix.

individuals, as is beautifully explained by Rousseau with respect to the main characters of his novel *La nouvelle Heloïse*:

> I conjured up love and friendship, the two idols of my heart, under the most ravishing images. I amused myself by giving them all the charms of the sex that I had always adored. I imagined two female friends rather that two of my own sex, because although such friendships are less common, they are more pleasing. I endowed my heroines with two personalities, different but matching, with two faces, not perfectly beautiful, but in accord with my own taste, and animated with benevolence and sensibility. I made one dark the other fair, one lively the other languid, one wise the other weak...[9]

In general, the idea is that what I called a character (or, more exactly, what I called here a character *stricto sensu*) is a set of salient properties. So far, we have seen how such an idea can account for situations like those illustrated by (8) or by the second part of (12b), where a character, conceived of as a type, that is a set of properties, is seen from the point of view of its external vicissitudes. The next, necessary step, is to make this notion more definite by introducing a suitable semantic framework. Yet, before addressing this issue, I have to mention a problem raised by other "complex" types of sentences about fiction. Consider the following sentences:

(16)   Charlus is taller than Danny DeVito
(17)   Charlus is taller than Sancho Panza.

Both sentences are intuitively true (for we know from the *Recherche* that Charlus is very tall, whilst we know from *Don Quixote* that Sancho Panza is very short; moreover, it is a fact that Danny DeVito is very short, too). The problem is that they seem to have the same structure as a sentence like:

(18)   Gerard Depardieu is taller than Danny DeVito

where two real persons are involved. But in (16) the comparison is between a character and a person, and in (17) between two characters. On the other hand, these characters are not generated by the same text, so that (17) cannot be assimilated to simple paratextual sentences such as:

(19)   Charlus is taller than Bergotte

which is not problematic because, intuitively speaking, we are referring to the world of the *Recherche* and we are speaking of properties that Charlus and Bergotte, as persons, turn out to have in *that* world. But

---

[9]This passage of the *Confessions* is quoted in Cranston (1991, p. 32).

such a reference is not possible in the case of (16) and (17). No text describes a state of affairs in which *both* Charlus and Sancho Panza (or Charlus and Danny DeVito) are persons: this is why, as I have already emphasized, we are inclined to say that the comparison, here, is between two characters (or between a character and a person). If characters, obviously, are not persons, how is it possible to account for the fact that we can say, truthfully, that a character x is taller than a person y or that a character x is taller than a character z (where x and z do not belong to the same story)?

10.4   Characters (*stricto sensu*)

I have already mentioned Frege's remarks about the occurrences of proper names in fictional contexts. In spite of their very general character, these remarks are based on a precise idea, which sounds quite intuitive. The idea is that, unlike the sentences of the ordinary discourse, the sentences occurring in a fictional context (i.e. *textual* sentences, in the terminology adopted here) do not determine genuine assertions: they are not used, in Frege's words, to state anything true or false. If this is so, Frege goes on, it is pointless to ask ourselves *what* is denoted by proper names like *Ulysses* (or, in our examples, *Vinteuil*, *Combray*, and so on) when they occur in these sentences.

Frege's remarks stop here. But if they are correct, we can wonder what the role of proper names (or of other singular terms) is in this case. We can wonder, for example, what the role of the proper names *Vinteuil* and *Combray* is in a textual sentence discussed at the outset:

(6)   M. Vinteuil s'était retiré auprès de Combray

In fact, one might object that, after all, even in these sentences the use of proper names allows for the attribution of properties and relations, exactly as in ordinary sentences like:

(20)   Fellini moved to Rome about sixty years ago.

A possible answer to this question along the lines of Frege's remarks is the following. There is no need to postulate any denotation for a proper name occurring in sentences like (6) because such a name is, here, a mere placeholder which the properties and relations at issue can be appended to. As the story develops, the agglomerate of properties and relations associated with such placeholders increases, thanks to sentences like (6).

I will not go into the details of an appropriate formal semantics, but the general idea is that at the end of this "storing" process what we get, in correspondence with a proper name like *Vinteuil* (or a definite

224 / Andrea Bonomi

description like *The red-haired woman who fights with Korablièva in a cell of the prison*[10]), is a set of characterizing properties and relations. This is what we call a *character* (*stricto sensu*) and since a peculiar feature of a character is its indeterminacy with respect to most properties and relations, we can also speak of a *type*: a type of person, a type of dog, a type of river, and so on. A character (*stricto sensu*) is what we can refer to in metatextual sentences. For example, we might say that the type (or character) Vinteuil has a particular property, like being artistically impeccable, or a particular relation, as being created by Proust or being loved by Orson Welles. Of course, the existence of characters is strictly related to the existence of the relevant texts (or stories[11]). This is why stories play a crucial role in the following definition of characters.

Let H be a story (e. g. Proust's *Recherche* or Tolstoj's *Resurrection*) and α a singular term (or, more exactly, a proper name like *Combray* or a definite description like *The red-haired woman who fights with Korablièva in a cell of the prison*). What must be specified is the theoretical meaning, in the present reconstruction, of the statement that α is a character of H.:

(21)   CHAR(α,H).

As we have seen, the intuitive idea is that whilst α is, *within* a story H, a person (or a town, a river, an animal, and so on), when considered from *outside*, i.e. as an object our statements are about, it is a *type*, or, more exactly, a set of properties and relations fixed by H. From this point of view, characters *exist* in our world (and, as a consequence, are not strange creatures like "non-existing" entities). And this is so because there exist, in our world, things like stories or texts, which make characters possible and which, in turn, are generated by some specific activity like writing or story-telling. As other cultural artifacts, e.g. numbers or symphonies, characters exist as abstract objects, whose existence depends on the existence of the relevant story. In this sense, as we will see later on, even the characters associated with real persons (e.g. Napoleon) are abstract entities created in connection with a story (e.g. what we often call *the Napoleon of War and Peace*). In general, the α of H (or, more succinctly, α_H) can be defined in the following terms:

(22) $\alpha_H = \lambda P[P(\alpha)]_H$.

---

[10]No proper name is associated to this character in Tolstoj's *Resurrection*.

[11]In what follows I will speak of a story and of the text in which that story is narrated without introducing the necessary distinctions. This simplification, which would be misleading in different contexts, is not relevant in the present discussion.

That is, a character α_H,[12] relative to a text or a story H, is the set of (relevant)[13] properties P such that in the story H we have that α is P.

For instance, the Combray of the *Recherche* can be characterized as follows:

(23) $\text{Combray}_R = \lambda P[P(\text{Combray})]_R.$

Notice that, in a definition such as (23), the proper name *Combray* occurs in the definiens *within* the scope of the index R (referring to the *Recherche*), which can be assimilated to an intensional operator. In other terms, the proper name is used to define the character in a purely *de dicto* way. Once more, the idea is that the Combray of the *Recherche*, which can be an object of discourse in our everyday language, is a type: i.e. the set of (relevant) properties and relations which, *in the text*, are associated to this proper name (seen, as suggested by Frege's analysis, as a mere placeholder). *Within* the story, of course, we do not have characters but persons (like Vinteuil), towns (like Combray), rivers (like the Vivonne), and so on: in short, we have ordinary individuals of familiar types (i.e. persons, towns, rivers, and so on). But assuming that in a story H there is a person x with such and such properties is quite different, of course, from assuming that there is a person x which in the story H has such and such properties. *Outside* the story, there is no individual of any familiar type (like a person or a town) we can refer to, but only a character, that is a *type* of individual. In a sense, a character, seen as a set of properties, is what we are left with when we try to bring a fictional individual out from its fictional milieu.

This kind of analysis, based on the idea that only general or (as I will say) generic statements are possible in the case of fictional entities like characters, can shed light on the problems raised by a sentence such as the one about Charlus and Sancho Panza, repeated here as (24):

(24)   Charlus is taller than Sancho Panza.

As I suggested, the difficulty is that on the one hand we seem to speak of characters as persons (see the analogy with a sentence like *Gerard Depardieu is taller than Danny DeVito*), but on the other hand Charlus and Sancho Panza are not persons, and there is no story H such that both of them are persons in H (whilst there is a story in which, for instance, both Charlus and Bergotte are persons: this is why a paratextual sentence like *Charlus is taller than Bergotte* states

---

[12]The index H will be omitted when the context is clear enough.

[13]I will not discuss here the problem whether a character α is defined by all the properties ascribed to it by the story H or by a subset of relevant properties.

something true of this two persons *with respect to the world of the Recherche*).

Fortunately, treating characters as types allows for a natural solution to the problem. Let us say that an individual x *instantiates* a character α (defined as before) if it satisfies all the salient properties in α. So, the meaning of a sentence such as (24) can be reconstructed in terms of generic sentences about types. More exactly, if we assume that the generic operator ranges over individuals and that it applies to a restrictive clause and a matrix,[14] what we get is something like:

(25) $\text{Gen}_{x,y}$ [Inst(x,Charlus) & Inst(y,Sancho Panza)] $\supset$ [Taller(x,y)]

or, if the generic operator ranges over situations too:

(26) $\text{Gen}_{s,x,y}$ [in s: Inst(x,Charlus) & Inst(y,Sancho Panza)] $\supset$ [in s: Taller(x,y)].

The intuitive idea is that the type "Charlus" and the type "Sancho Panza" are such that, in general, anyone who has the characteristics associated with the former type is taller than anyone who has the characteristics associated with the latter type. In other terms, a sentence like (24) is assimilated to traditional generic sentences like *Rabbits are taller than rats* or *A Ferrari is faster than a Maserati.*

10.5   The principle of importation

We can go back to function g that, as we have seen, allows us to characterize the relation which in some cases holds between individuals (like persons or towns in the actual world) and those particular cultural artifacts that we have called characters (in a broad sense of the term). Of course, not every character is the relatum of this kind of relation. Take, for instance, Sherlock Holmes or Lilliput. As far as we know, it is reasonable to think that there is no real person α such that $g(\alpha, D, X)$ = Sherlock Holmes, no real town β such that $g(\beta, S, Y)$ = Lilliput, where D and S are the relevant stories and X and Y are the sets of relevant properties. But in other cases such an entity does exist. For instance, the relation between Combray and its real counterpart, Illiers, is so strong that the latter was officially renamed Illiers-Combray in virtue of a legislative decree signed by the President of the French Republic. There are also situations in which, if two different sets X and Y of relevant properties are selected, two different individuals can be

---

[14]Roughly speaking, 'Gen_x $[A(x)]$ $[B(x)]$' means something like 'In general it is true that if x satisfies A, it also satisfies B'. See Carlson and Pelletier (1995) for the semantics of the generic operator.

associated to the same character. This is the case, for instance, of *The Profession of Faith of the Savoyard Priest*: as Rousseau himself reveals in the *Confessions*,[15] the Savoyard Priest, as a fictional character, can be seen as the counterpart of two different people on the basis of different sets of characterizing properties. Such a situation might described in terms of a "double" identity:

(27) $g(\text{Gâtier}, D, X) = $ The Savoyard Priest
(28) $g(\text{Gaime}, D, Y) = $ The Savoyard Priest

But it is time to give some conceptual substance to function g, which we refer to in order to express this kind of relation. To do this I will resort to a more accurate version of the principle of importation:

> (IMP) Let X be the set of salient properties which are assumed to characterize α in the background information, and let Y be a particular subset of X. Then:
> (a) $g(\alpha, H, Y) = \beta$ iff, for every property P in Y, $P\alpha \rightarrow (P\beta)_H$
> (b) for every property P in X, $P\alpha \Rightarrow (P\beta)_H$

where '$(A)_H$' means, as before, that in the context of the story H we have that A, whilst the double arrow denotes a default entailment. So, (a) says that β is the character, in H, corresponding to α if and only if β satisfies, in H, a set Y of salient properties which are *selected* among the properties characterizing α. In other words, (a) defines the correspondence, for instance, between a (real) person and a character in terms of the salient properties that we are willing to transfer from the former to the latter.[16]

Moreover, as stated by (b), when P is known as a salient property which characterizes α, P is also one of the properties of β in H, *unless otherwise stated in H*. The intuitive role of function g and its "converse"[17] g^* can be illustrated by an example. After reading Painter's

---

[15]Book III: "By putting together Monsieur Gâtier and Monsieur Gaime I made the original of the Savoyard Priest out of these two respectable priests".

[16]The nature of this kind of information can obviously vary according to the context. In many cases it can involve a *causal* relation between the "model" (Illiers) and the author (Proust); it can also involve some stereotypical properties, or the *author's intentions*, and so on. I will not address this problem here. It should also be noticed that this kind of information is sometimes restricted to a small circle of experts. For instance, in the case of the *Recherche* the common reader may not be aware of the relation between Illiers and Combray, which is well-known in the circle of Proustian scholars.

[17]By the converse of a function g such that $g(\alpha, H, X) = \beta$ I mean the function g^* such that $g^*(\beta, H, X) = \alpha$.

book on Proust, I can say[18]:

(29)   Illiers, in the *Recherche*, is Combray

or, more explicitly,

(30)   The Illiers of the *Recherche* is Combray.

Similarly, I can say:

(31)   Combray, in the real world, is Illiers

or, in the more complex situation described above:

(32)   The Savoyard Priest, in the real world, is Gâtier. But, from a different point of view, he is also Gaime.

In the present framework identities of this kind, which are intuitively true, can be accounted for by means of function g (and its converse), as characterized in (IMP). In fact, this approach allows us to explain the nature of the relation between the individual denoted by the proper name *Illiers* (which is a little town near Chartres) and a character associated with the *Recherche*. The relevant identity sentence is:

(33) $g(\text{Illiers}) = \text{Combray}$,

or, more explicitly:

(34) $g(\text{Illiers,R,X}) = \text{Combray}_R$

where g is the function which, on the basis of the story R (the *Recherche*) and considering the set of relevant properties X provided by the background information, maps individuals, like Illiers, to characters, like Combray, generated by the text. To simplify things, if the situation is clear enough, this kind of identity will be expressed by sentences like (33) rather than by more explicit sentences like (34).

So, thanks to the principle of importation the meaning of (29) is expressed by a metatextual sentence like (33), whilst the meaning of (31) is expressed by this other sentence:

(35) $g^*(\text{Combray, R,X}) = \text{Illiers}$.

Notice that, thanks to g, it is also possible to explain why proper names like Combray, in their use *outside* Proust's text, can have a

---

[18]Fauconnier (1985) discusses some similar sentences. For example:

In the picture, Lisa is the girl with the blue eyes.

What is peculiar to our examples is that the apparent identity statement involves two proper names. (In this connection, Fauconnier presents other interesting examples, but of a different type with respect to (29). For instance: In the movie, Liz Taylor is Cleopatra).

double reading. Consider, for instance, the following sentences:

(36)   In the real world Combray does not exist
(37)   In the real world Combray exists (and it is Illiers).

Both of them can be true. In fact, the former can mean that there is no real town whose name is *Combray*. This reading can be accounted for by the following formula:

(36') $\neg \exists x (\text{town}(x) \ \& \ x \ = \ \text{Combray})$

The latter can mean: there is a real town whose counterpart in the *Recherche* is Combray (and this town is Illiers). This reading can be represented as follows:

(37') $\exists x (\text{town}(x) \ \& \ g(x) = \text{Combray})$

10.6   Ordinary proper names

Function g, as we have just seen, associates a character (if any) to an individual (like Illiers, for instance). So, in the case of sentences such as (29) it would be inappropriate to speak of an identity in the logical sense of the term. The idea is that (29) is to be accounted for by means of a formula like (33), where, strictly speaking, what is stated is not the existence of an identity relation between the Illiers of the real world and the Combray of the *Recherche*, but the existence of a correspondence between these two entities, as expressed by function g. This is quite intuitive, for Illiers is a real town and Combray a character generated by a text, and it would be absurd to maintain that they are the same thing. In fact, (29) can also be paraphrased as follows:

(38)   In the *Recherche* Illiers becomes Combray

where the *is* of the apparent identity has disappeared. In short, in such cases the value of function g (which is a character) does not coincide with its argument (which is a town).

This means that sentences such as (29) must be distinguished from ordinary identity statements like:

(39)   The evening star is the morning star
(40)   Hesperus is Phosphorus

where only one entity is involved (i.e. Venus). In fact, unlike identity, g does not coincide with its converse – i.e. the function which maps characters to individuals – which is referred to in sentences like (31),

repeated here:

(31)   Combray, in the real world, is Illiers

or, more explicitly,

(41)   The Combray of the real world is Illiers.

In this case it is quite evident that to exchange the designators *Combray* and *Illiers* (as we can safely do in genuine identity statements like (40)) would give rise to an odd statement or, in any case, to a sentence whose meaning is not equivalent to the meaning of (31) (as suggested by the different roles that the phrase 'of the real world' would play in the two sentences):

(42)   The Illiers of the real world is Combray.

This is why, in the cases we have taken into consideration so far, it is reasonable to maintain that sentences like (29) or (30) do not state the identity between the referents of the designators (e.g. *Illiers* and *Combray*) involved in those sentences, but only a correspondence between a town and a character. And this is correctly mirrored by g, which is not the identity function.

In general, as we saw in the last section, the identity

(43) $g(\alpha, H, X) = \beta$

expresses a relation between an individual α and a character β generated by H (the story, or the text): in our example, since the proper name at issue is *Combray* and the story is the *Recherche*, this character can be referred to by the complex term *The Combray of the Recherche*. (Other examples of terms denoting characters are: *The Buck Mulligan of Ulysses*, *The aggressive red-haired woman of Resurrection*, *The Savoyard Priest of the Confessions*, and so on.)

So far, so good. We can easily admit that, as shown in the above examples, there is a clear sense in which a character (like the Combray of the *Recherche*) is distinguishable from an ordinary individual (like Illiers), even though there is a very close relation between that character and that individual. Notice, however, that in the examples we have discussed so far *two* names are involved: one for the character (e.g. Combray) and one for the corresponding individual (e.g. Illiers). But what about proper names such as *Paris*? In this case we have a *single* name which occurs both in factual sentences of our everyday life and in fictional sentences, for instance in the *Recherche*. And in both cases the name is supposed to refer to the same thing, that is Paris. As a consequence, we can wonder whether the distinction between the city (Paris) and the character (the Paris of the *Recherche*) still makes

sense. A negative answer might be based on the following argument.

When we find an occurrence of a common noun like *dog* or *tree* in the *Recherche*, it would be quite absurd to assume that Proust uses these words in some "special" sense. Whatever the meaning of a common noun may be, *dog* refers to dogs and *tree* refers to trees, and there is no reason to question such truisms when we read the *Recherche* or any other novel. The same holds of proper names: whatever the meaning of the name *Paris* may be, this name refers to Paris, and that's all. Whilst Combray does not exist (so that it is reasonable to keep this "character" distinct from Illiers, which does exist), Paris exists: this is why one might conclude that there is no need, this time, to keep the character (the Paris of the *Recherche*) distinct from the real town. Exactly as the words *dog* and *tree* preserve, in the text, their usual meaning, the word *Paris* preserves its usual reference, that is Paris itself, and to speak of two entities (the city and the character) is misleading.

In spite of its apparent cogency, this argument is based on a double misunderstanding. First of all, it should be noticed that speaking of characters in connection with proper names like *Vinteuil*, *Combray* or *Paris* does not entail that such names are intended to denote (or refer to) characters when they occur in the *Recherche*, i.e. in *textual* sentences. Characters, as theoretical entities, are introduced to account for the occurrence of singular terms in statements that *we* can make *about* a story and its characters (that is metatextual statements, in the terminology adopted here). But this has nothing to do with the truism that a name such as *Paris*, when it occurs in the *Recherche*, has no special meaning, no special reference. Indeed, it can easily be granted that this name preserves, in the *Recherche*, its usual reference. This truism, however, is perfectly consistent with the idea that, in our talks about fiction, we can speak of a character associated to a proper name like *Paris*. Moreover, once we have admitted that in fictional contexts ordinary proper names preserve their usual referents, it is still possible to observe something very peculiar in this fictional use of ordinary proper names (with their usual referents). This peculiarity concerns the relationship between the name and the logical space in which it is located.

In ordinary contexts, e.g. in sentences reporting factual events, the use of an ordinary proper name like *Paris* or *Kutùsov* entails a particular rearrangement of the logical space in which their referents (a town and a person, respectively) are represented. But this is not what happens when such names are used, for instance, in a novel. Consider the name *Dreyfus*, which also occurs in the *Recherche*. If we read in a history book that Dreyfus received a sympathetic letter from C. Debussy,

the description of this event allows us to update the set of properties and relations which are to be considered as characterizing Dreyfus. This description is stored as part of the information concerning the intended individual. On the contrary, if we read in the *Recherche* that Dreyfus received a sympathetic letter from Vinteuil, such an updating would not be justified: the event at issue is not stored as part of the information concerning Dreyfus that should be added to the common ground and, in this connection, the fact that the name *Dreyfus*, in the *Recherche*, denotes Dreyfus is not relevant. What counts is that, in terms of logical spaces, the use of such a name in fictional contexts cannot be identified with its "ordinary" uses. Let us see why.

10.7 Logical spaces

What characterizes the sentences occurring in a fictional text is the possibility of canceling assumptions that are part of the relevant background information, i.e. a body of information that is presumed to be available to the participants in the communicative exchange. In the second clause of principle (IMP) this peculiarity is accounted for by the type of entailment used to qualify the relationship between what an agent assumes to be the common ground and the part of it that, according to this agent, should be imported in the story. The idea is that this entailment relation is not the classical one (which would mean that any assumption in the relevant background information should be imported in the story) but a *default* entailment relation: any assumption in the given background information holds in the story *unless otherwise specified by the story itself*. For example, suppose that nowhere, in the *Recherche*, is it explicitly stated that Paris is the capital of France. Well, there is no problem in attributing such a property to the Paris of the *Recherche* in virtue of the general assumptions concerning that city, provided that nothing in the text suggests that Paris is not the capital of France.

The defeasibility, in a fictional story, of the assumptions in the given background information (as suggested by the default entailment) is crucial to understand the reason why one often says that, unlike ordinary declarative sentences, textual sentences do *not* give rise to genuine assertions. Let us address this problem.

Let **X** be the information that the agent presumes to be the common ground, that is the set of propositions whose truth is taken for granted (in relation to a given object of discourse). From a formal point of view, **X** can be seen as a *logical space* **B**, i. e. the set of situations or possible worlds which are compatible with those propositions. More precisely,

if a proposition is considered as a set of situations (that is, the set of situations in which the proposition is true), the logical space $\mathbf{B}$ can be defined as follows:

(44) $\quad \mathbf{B} = \{w \in W \colon w \in p$ for every proposition p in the (presumed) common ground $\mathbf{X}\}$

That is, the logical space associated to the presumed common ground $\mathbf{X}$ consists of the situations where all the propositions in $\mathbf{X}$ are true.

Intuitively speaking, $\mathbf{B}$ is the set of "living options" selected by the background information[19] and what we intend to do, when we make an assertion, is to restrict the set of these options, not to destroy it (unless we want to question $\mathbf{B}$ itself: but in the present discussion we will ignore such situations). In particular, an obvious assumption concerning $\mathbf{B}$ is that: (i) $\mathbf{B}$ should contain w_0 (the actual world); (ii) the incoming information should allow us to eliminate from $\mathbf{B}$ only counterfactual worlds, not w_0. In short, the idea is that, in principle, $\mathbf{B}$ is intended to contain only *true* information. It is in this sense that we say that $\mathbf{B}$ is w_0-oriented.

When a sentence A is uttered in the context $\mathbf{B}$, this utterance determines an updating of $\mathbf{B}$, in the sense that all the situations which are not compatible with the proposition expressed by A (in that context) are eliminated. But let us consider what happens when a sentence such as:

(45)   Leo, a stammering Afghan snake-charmer, moved to Rome

occurs in a fictional text. If, as before, $\mathbf{B}$ is the information, concerning Rome, that is presumed to be the relevant common ground, reading (45) in the novel does not lead us to restrict $\mathbf{B}$ by eliminating all the situations in which no Afghan snake-charmer moved to Rome. Such a modification of $\mathbf{B}$ is justified in the case of (20) (i.e. the sentence 'Fellini moved to Rome about sixty years ago'), whose occurrence in a report about Rome determines the elimination, from $\mathbf{B}$, of all the possible situations where Fellini did not move to Rome. But the occurrence of (45) in a novel does not determine a similar modification of $\mathbf{B}$: that no stammering Afghan snake-charmer has ever moved to Rome is still a live option in $\mathbf{B}$. Unlike the occurrence of the proper name *Rome* in

---

[19]This characterization of the background information is very close to the notion of context set defined in Stalnaker (1999). Perhaps a more appropriate theoretical framework would be a multi-agent analysis (see Arló-Costa's paper in this volume), where what is crucial is not the body of shared assumptions as such, but the speaker's (hearer's) beliefs about these assumptions. Yet, the problems I address in the present paper are independent of this issue. This is why I resorted to a familiar model like Stalnaker's.

(20), the occurrence of this name in (45) is *not* anchored to that logical space.

More intuitively, saying that a genuine assertion like (20) restricts the given logical space $\mathbf{B}$ is tantamount to saying that, thanks to (20), we assert something new *about* Rome. From this point of view it is possible to explain why, as one often says, textual sentences such as (45) are not used to make genuine assertions. The idea is that in this case there is no contraction of the relevant logical space, i.e. no real increase in the background information.

In general, what happens can be summed up as follows. When the utterance of a sentence A is used to make a genuine assertion, the relevant context $\mathbf{B}$ can be assimilated to a logical space which fixes the boundaries within which it is possible to identify the set of alternatives selected by this utterance. As regards the occurrence of A in a fictional text, the situation is different, since $\mathbf{B}$ must be replaced by a different logical space $\mathbf{B}_H$ and a suitable *revision* of $\mathbf{B}$ is required to fix this new context. To be sure, the principle of importation tells us that, in virtue of a default entailment, $\mathbf{B}_H$ will preserve several salient features characterizing $\mathbf{B}$, because $\mathbf{B}_H$ is obtained by a "conservative" revision of $\mathbf{B}$. But the moral that we can draw from the discussion of (45) is that the appropriate interpretation of such a textual sentence is crucially based on this kind of context shift[20].

10.8 Occurrence vs. scene-setting

We have just seen that in textual sentences a proper name is not used to make genuine assertions (with respect to the presumed common ground). So we can ask ourselves whether such an approach can shed some light on our main problem, which concerns the relation of *scene-setting*. In fact, we have introduced this concept because, strictly speaking, a fictional event, described by a textual sentence or a concatenation of textual sentences, cannot have an occurrence relation with a real place or time in our world. The problem is that sentences like (46) or (47) seem to make sense:

(46) The events described by Gadda's *Pasticciaccio* take place in Rome

(47) The murder of Mrs Balducci takes place in Rome

whilst in my analysis *only* real events or states can be in the occurrence

---

[20]This point is discussed in the Appendix in connection with paratextual sentences.

relation with Rome. This is why whilst a sentence such as:

(48)   The murder of Matteotti took place in Rome

can be associated to a logical form like:

(48')   Occ(e,Rome),

where e denotes a real event (i.e. the murder of Matteotti), (47) cannot.

If this is so, I have to specify how sentences like (47) are accounted for in the theoretical framework under discussion. As anticipated at the outset, the idea is that the relation of occurrence should be replaced by the relation of *scene-setting*. Although it is true that the event at issue (that is the murder of Mrs Balducci) has a spatial extension in the world of the *Pasticciaccio*, according to the downward indeterminacy argument it has no extension in our world. As a consequence, it is impossible, in principle, to assign to this event an extension in the region of the physical space around us occupied by Rome. Yet, it is reasonable to maintain that Rome inspired Gadda when he had to build up the "scene-setting" for the events described in his novel, and in particular the murder of Mrs Balducci. So, if e is this event and G is Gadda's text or story, we would have something like this:

(47')   SSET(e,Rome,G)

where SSET is a three-place relation between the event[21] at issue, the city and the story. By referring to function g, in general this scene-setting relation can be characterized in the following terms:

(49)   SSET(x,y,H) iff, for some character α_H,
         $g(y) \; = \; \alpha_H$ & $\lambda v \; [\text{Occ}(x,v)] \in \alpha_H$.

In other words, the relation of scene-setting holds between an event x, a certain place (or time) y and a given story H if and only if in H there is a character α corresponding to y such that the property of being a place where x occurs belongs to α: that is, if and only if there is a character α corresponding to y such that, in H, x and α are in the relation of occurrence.

Appendix: A note on the semantics of indices

The role of the prefix. I have spoken of indices as a useful device, in logical forms, to account for the reference to the intended context of discourse, such as a novel or a film. All this is quite generic, of course, and some qualifications are in order.

---

[21]It would be more appropriate to speak of a *type* of event, but I will not address this problem here.

First of all, it should be specified that a context, in the present reconstruction, is not a simple package of relevant parameters such as the time (the place, the agent, etc.) of an utterance. For the reasons that I discussed in a previous section in connection with paratextual sentences, a context should be assimilated here to a body of information that is presumed to be available to the participants in the communicative exchange, and the assumptions about the time (the place, the agent, etc.) of the utterance are part of this presumed common ground. In this sense, a context can be seen as a set of propositions or, in a suitable framework, as a set of situations: the situations compatible with the information which is presumed to be shared by the agents (including the information concerning the current utterance). This is why, in what follows, a "prefix" such as *In the novel N* or *In the movie M* should be considered as a context shifter, that is an operator which makes a particular context relevant to fixing the content of the sentence to which the prefix is applied and to evaluating it as true or false.

To see why we should speak of a *context shifter* and not simply of a world shifter (as in most of the classical approaches) some new examples are in order.

Consider for instance the following sentences:

(1)   Napoleon Bonaparte is an arrogant person
(2)   John told me that Mary is pregnant
(3)   In *War and Peace* Napoleon Bonaparte is an arrogant person.

Interestingly enough, the use of the present tense sounds quite natural in the case of (3), whilst using a past tense (namely *was* instead of *is*) would sound odd. By contrast, in the case of (1) the opposite is true: normally, the use of a past tense is much more natural. Moreover, the use of the present tense in (3), unlike its use in (2), does *not* entail that the eventuality at issue (i. e. Napoleon's being an arrogant person) is a present eventuality. Whilst (2), in its natural interpretation, entails that the utterance time is included in the time interval corresponding to Mary's pregnancy, a similar entailment is not allowed in the case of (3). This is tantamount to saying that the present tense, in (3), is not to be interpreted with reference to the utterance time. In general, since tenses are indexical elements whose denotations depend on the context, a plausible explanation of the peculiarity of (3) is that one of the effects of a prefix such as *In War and Peace* is to determine, among other things, a new context for the interpretation of the present tense. (See Zucchi (2001) for a formal treatment of this phenomenon.)

As for space, consider the opposition between expressions such as

to come/to go or to *to be behind/to be in front* which are often mentioned as indexical expressions (e.g. in Chierchia and McConnell-Ginet (1992)). Take, for instance, sentence (4). If I am in Milan, only the first option is acceptable (in normal situations), whilst the second would sound very odd:

(4)   Yesterday, Leo came (? went) to Milan to deliver a letter.

On the contrary, *both* options are available in the case of paratextual sentences, as witnessed by the acceptability of both variants (even if the speaker is in Milan):

(5a)   In the *Promessi Sposi*, Lorenzo Tramaglino comes to Milan to deliver a letter
(5b)   In the *Promessi Sposi*, Lorenzo Tramaglino goes to Milan to deliver a letter.

Since a similar argument holds for the opposition *to be behind/to be in front*, a natural explanation of this phenomenon is that sentences such as (5b) are acceptable because, in general, the prefix *In fiction F* makes a new point of view relevant: a point of view which is determined by a context shift, for the location of the actual speaker is no longer relevant.

Finally, consider this other sentence:

(6)   ?John is always a very erudite person.

In most cases, a sentence of this kind would sound very odd. But there are situations in which it is perfectly acceptable. If, for instance, we are speaking of the different versions of the Faust legend, a sentence like:

(7)   Faust is always a very erudite person

is not problematic in the interpretation which can be paraphrased by the conjunction: In Marlowe's tragedy Faust is a very erudite person and in Goethe's tragedy Faust is a very erudite person and ... So, a possible explanation of the acceptability of (7) is that the variable bound by the adverb of quantification *always* is a variable over contexts: what we are considering is the set of different backgrounds of information against which this character appears. If C is the set of relevant contexts we are referring to (i.e. Marlowe's version of the tragedy, Goethe's version, and so on), (7) can be associated with a quantificational structure like:

(8)   For every c in C, in c Faust is a very erudite person.

(It should be noticed that taking C as a set of more familiar entities

such as possible worlds would be problematic here. Indeed, we cannot refer to the set of possible worlds which are compatible with the story (as is usual in such cases) for the simple reason that that there is no single story here, but a set of stories which are inconsistent with each other. The least we can say, if we want to preserve the possible world machinery, and if contexts are formalized as sets of possible worlds, is that C is a set of sets of possible worlds.)

Indices, worlds and contexts. Once we have recognized, for independent reasons, the need for variables ranging over contexts, there are two points that deserve a more detailed discussion.

Indices, as I have characterized them in the present paper, can be seen as variables over contexts which occur in logical forms. These variables are introduced to account for prefixes of the type *In the novel N* in paratextual sentences such as:

(9)   In the *Recherche*, Vinteuil's daughter likes dancing with girls

Yet, it should be noticed that, in many circumstances, if we want to express the same content we do not use "prefixed" sentences like (9), but a simpler sentence like:

(10)   Vinteuil's daughter likes dancing with girls.

So, a first question is: (i) *what is the exact relation between sentences like* (9) *and* (10)?

In the present discussion, when speaking generically of contexts, different interpretations are possible, because in some cases we mean by context what is relevant in order to determine the content of a sentence, whilst in other cases we mean by context what is relevant to evaluate this content. This time, then, the question is: (ii) *are indices, that is variables over contexts, relevant to accounting for both notions of context*?

To answer question (i) we can start from a very intuitive remark: when the reference to the story in question can be taken for granted (when, for instance, I am speaking with a friend of mine who has a copy of the *Recherche* in her hands and who is asking me about Miss Vinteuil's habits), using a sentence like (10) is quite natural; on the contrary, (9) would sound pedantic (because it conveys redundant information) and, as such, unnatural. The idea is that the story is, here, the relevant context of discourse in the sense that *it provides us with the information we need* to perform two essential tasks: (A) to determine the content of the sentence at issue; (B) to evaluate this sentence as true or false. With respect to (A), the *Recherche*, in the above example, can be seen as the source of the background information we

presuppose. Without this kind of information we would be unable to grasp the content of a sentence such as (10): for example, we would be unable to assign a reasonable interpretation to the name *Vinteuil*, to the definite description *Vinteuil's daughter*, or to the use of tenses and other indexical expressions. Yet, this is just half the story, for the information provided by the *Recherche* is relevant in another crucial aspect: it selects the world[22] with respect to which the sentence is to be evaluated. It is in this sense that we say, intuitively, that a sentence such as (10) is true "in the world" of the *Recherche*.

We have just considered a case in which a "prefixed" sentence like (9) sounds unnatural, whilst its "unprefixed" counterpart sounds quite appropriate. As we have remarked, the oddity of the former type of sentence is due to the fact that, since the suitable reference to the relevant information is contextually *given*, using the prefix *In the Recherche* would be redundant. For symmetrical reasons we must expect that, when the context of the discourse (or, more exactly, the background information which determines the content of the sentence and the world of evaluation) is *shifted*, this fact should be *signaled*, and the prefix is the appropriate tool to do that. Indeed, this is what happens in the following examples:

(11)   The word *Rosebud* is the name of a friend of mine.
       It is the name of a sled, too

(12)   The word *Rosebud* is the name of a friend of mine.
       In *Citizen Kane* it is the name of a sled, too.

Using an unprefixed sentence such as (11) might be misleading in normal circumstances, because, ceteris paribus, a hearer (or a reader) who does not know Welles' movie might be led to think that, since in the first part of the discourse we are speaking of real individuals, there is some real sled whose name is *Rosebud*. So, to avoid this misunderstanding, the prefixed sentence is, this time, quite appropriate, as shown by (12): the presence of the prefix *In Citizen Kane* signals the shift of the relevant background information, which is no longer what we assume to be known with respect to the world around us, but what we assume to be known with respect to a particular story.

As a conclusion, a first answer to question (i) might be condensed in the following remark: in the situations we have considered above, what makes either a prefixed sentence or an unprefixed one appropriate is just a pragmatic factor. Where the reference to the relevant background information (originated by a story) is taken for granted, the prefix is

---

[22]To simplify things, in what follows I will often speak of a single world, instead of a plurality of worlds, associated with a given story intended as a context.

redundant and using it would be pedantic, but if such a reference is not obvious, or if it is shifted, the prefix is quite appropriate.

There is something vague, of course, in this way of speaking about the role of context in making prefixed or unprefixed sentences more appropriate, but, in a sense, this vagueness is a necessary characteristic of such an analysis, due to the variety of the situations we have to deal with. For instance, I have just said that in many circumstances (12) is more appropriate than (11) to express, among other things, a crucial fact in Welles' movie. But if the sentence should be uttered during a conversation between two movie critics, the prefix might be redundant and it might be cancelled (or even replaced by a phrase like *As everybody knows...*). In other words, in such a situation it is probable that (11), rather than (12), would turn out to be more appropriate.

To consider another example, take the following pair of sentences:

(13)   The Sultan of Congo likes jokes
(14)   In *Les bijoux indiscrets*, the Sultan of Congo likes jokes.

If uttered *ex abrupto* or during a conversation about sovereigns who are known to like jokes, (13), unlike (14), would be quite inappropriate; but, once more, it would sound perfectly natural in other circumstances: for example during an explicit discussion about the characters created by Diderot.

From a formal point of view, indices, as variables over contexts, are a useful tool to account for both types of paratextual sentences (i.e. prefixed and unprefixed sentences). I will not go into the details of the formalism and I will content myself with a sketchy presentation. As I have just recalled, indices can be seen as variables over contexts.[23] They are unpronounced items which have no observable counterpart in surface structures and can occupy different positions in logical forms. To simplify things, let us assume that such silent variables can be associated, respectively, with the Noun Phrase and the Verb Phrase (or, more exactly, with the clausal structure projected by the verb),[24] so

---

[23]The need for variables over worlds or situations has been motivated by several authors in connection with other issues. See Bonomi (1998) for the use of indices, as variables over situations, as an alternative, and more flexible, way of accounting for the "transparent" reading of intensional sentences: a reading which is usually accounted for by the "exportation" of the relevant material made possible by the lambda operator. A systematic treatment of variables over situations is presented in Percus (2002). The reader can refer to this paper for the technical details and the bibliography. In a sense, resorting to variables over *contexts* (rather than worlds) can be seen as a generalization of this kind of approach.

[24]Such a presentation is far from being accurate, but it is sufficient for the present

that we get structures like:

(15)   [...[... NP ...]$_c$ ... VP ...]$_c$

i.e., in our example

(13')   [[The Sultan of Congo]$_c$ likes jokes]$_c$.

It should be noticed that (13') contains free occurrences of the variable c. Therefore, the problem is: how can we obtain a suitable value for these occurrences of the variable? The idea is that in the case of (14) what is responsible for the binding of the variable over contexts is the prefix itself, which is conceived of as an operator:

(14')   (In *Les bijoux indiscrets*, c)
          [[The Sultan of Congo]$_c$ likes jokes]$_c$.

The effect of applying this operator to a structure of the type of (13') is to select the background information provided by the story. Since the context at issue is Diderot's novel, (14') is true iff the person who is the Sultan of Congo in this context has, in this context, the property of liking jokes.

So far, so good. But what about a sentence such as (13), which, unlike (14), is not characterized by the presence of the prefix? The answer is that if the reference to the story in question can be taken for granted (as in the original example), the context which must be selected as the value of the variable is easily found: it is the story itself. In other terms, what we have here is a sort of "indexical" binding of the silent variable, whose value is identified with the relevant context, that is the background information associated with the story. A little more exactly, such a situation is expressed by the following formula:

(13")   (In X, c) [[The Sultan of Congo]$_c$ likes jokes]$_c$

where X is contextually anchored to the story (*Les bijoux indiscrets*, in our example). The role of this indexical operator is to bind the context variable: that is, from an intuitive point of view, to specify what kind of context is relevant here. So, (13") provides us with the intended interpretation of (13') by fixing the relevant context: as in the case of (14'), (13") is true iff the person who is the Sultan of Congo in the context of *Les bijoux indiscrets* has, in this context, the property of liking jokes.

Opaque and transparent readings.   Before investigating more closely the interaction of a prefix such as *In the story S* with other intensional operators, let us reflect for a while on the nature of the silent variables

purposes.

I have just introduced to account for paratextual sentences. We have seen that, even in logical forms that correspond to simple sentences like (9) or (13) there are *two* places where indices can occur: one associated with the VP and the other with the NP.[25] In the above examples both positions are occupied by the same index. But they might be occupied by different indices. To see why this option is needed consider the following sentences:

(16)   In *Les bijoux indiscrets*, the Capital of Congo is inhabited by many gossipy persons

(17)   In *Les bijoux indiscrets*, the Capital of France is inhabited by many gossipy persons.

From an intuitive point of view, the main difference between (16) and (17) is that in the natural interpretation of the former sentence the NP *The Capital of Congo* has an "opaque" reading, whilst in the natural interpretation of the latter sentence the NP *The Capital of France* has a "transparent" reading. This is so because (16) is appropriate in a situation where we are speaking of the city which is the Capital of Congo *in* the context of Diderot's novel, i. e. from a point of view which is internal to the story. To do that we use the same definite description which is used in the novel. On the contrary, in (17) the way we refer to this city (which actually is Paris) mirrors *our* point of view, which is external to the story. (In Diderot's novel the proper names of the characters corresponding to Paris and France are, respectively, *Banza* and *Congo*). To see another illustration of the transparent reading consider the following example:

(18)   In *Les bijoux indiscrets*, some enemies of Diderot say ridiculous things.

Needless to say, the term *enemy of Diderot* does not occur in the novel, but it can be used, *by us*, to identify from outside some characters in the story. More exactly, in (18) the property of being an enemy of Diderot identifies some persons *in the context of empirical facts concerning our world*, and what this sentence means (in its natural interpretation) is that *these* persons say ridiculous things *in the context of the novel*. The opposition between the opaque reading and the transparent one can be reconstructed, in the present framework, in

---

[25]See Percus (2002) for the details (in particular, for the constraints that govern these variables).

terms of different indexing patterns:

(16') (In *Les bijoux indiscrets*, c)
       [[The Capital of Congo]$_c$ is inhabited . . .]$_c$
(17') (In *Les bijoux indiscrets*, c)
       [[The Capital of France]$_r$ is inhabited . . .]$_c$.

Here r is the index which is anchored to the background information presumed to be true by the speaker, whilst c is bound by the operator corresponding to the prefix. As desired, (16') means that the city which, in the context of Diderot's novel, is the Capital of Congo is inhabited, in that context, by many gossipy persons (opaque reading), whilst (17') means that the city which, in the context of empirical facts concerning our world, is the Capital of France, is inhabited, in the context of Diderot's novel, by many gossipy persons (transparent reading). In a similar way, the only natural interpretation of (18) is the one where the NP *Some enemies of Diderot* has the transparent reading, captured by a structure like:

(18') (In *Les bijoux indiscrets*, c)
       [[Some enemies of Diderot]$_r$ say ridiculous things]$_c$.

Notice that using this kind of structure to account for the opposition between the opaque reading and the transparent one provides us with an implicit answer to question (ii), that is the question about the different roles that indices can play. Indeed, indices have a double role here, for c is intended to select the context of evaluation, i. e. the (type of) situation in which the sentence following the prefix is to be evaluated as true or false, but r is intended to refer to a context (distinct from the context of evaluation) which fixes the reference of the noun phrase.

In the above analysis the prefix is treated as an intensional operator. To see how it interacts with other intensional operators and how the double role of indices I have just illustrated can explain some interesting structural ambiguities, consider this new example. Suppose that I am speaking with a friend about the cities where I would like to live and that I say:

(19) I wish I lived in the Capital of Congo.

Consider these three possible interpretations of (19): (i) the context c of my desires is such that in the scenario determined by c I live in the city which, in this scenario, is the Capital of Congo; (ii) the context c of my desires is such that in the scenario determined by c I live in the city which, in the context u of Diderot's novel, is the Capital of Congo; (iii) the context c of my desires is such that in the scenario determined by c I live in the city which, in the context r of empirical facts concerning

our world, is the Capital of Congo. A little more exactly, three different propositions (as sets of possible contexts) can be associated with my desires, respectively:

(i) λc[I live in [the Capital of Congo]$_c$]$_c$

(ii)λc[I live in [the Capital of Congo]$_u$]$_c$

(iii)λc[I live in [the Capital of Congo]$_r$]$_c$

Is the interpretation suggested in (ii) possible? To my intuition it is (provided that the reference to Diderot's novel is clear from the context): what we mean, on this reading, is that we would like to live in the city described by the novel. Such an interpretation is expressed by (ii), where the context u (i.e. *Les bijoux indiscrets*) is not the context with respect to which the embedded sentence should be true (for the simple reason that the context with respect to which it is true that I live in the Capital of Congo is not the context u of the story but the context c of my desires). On the contrary, u is here the context we refer to in order to get the intended interpretation of the NP *The Capital of Congo*, i. e. in order to interpret the definite description. This is tantamount to saying, once more, that the context associated with a story can intervene as an essential factor in the interpretation of a subpart of a sentence.

Context shift. In the present framework (whose full formalization will be the object of another paper) a context $c \in C$ is a set of propositions that are assumed to be the common background in a communicative exchange. As specified above, this set can also be identified with a set of possible situations: the set of situations in which these propositions are true.

> For every expression α, context c and situation w in c, $[\![\alpha]\!]_{c,w}$ is the content of α in w with respect to c. In particular, if φ is a sentence, $[\![\varphi]\!]_{c,w}$ is a function from c to 0, 1 such that, for any w'$\in$ c, $[\![\varphi]\!]_{c,w}$(w') $= 1$ iff w'$\in$ I(φ) (where I(φ) is the set of situations in which φ is true).

Crucially, besides variables, we must have names for contexts. For example, *War and Peace* (or some suitable constant) is such a name. Semantically, a possible solution is to treat these terms as rigid designators which denote the same content with respect to every context c (and every situation in c), that is:

> if S is a term denoting a context (e. g. a story like *War and Peace*), for every c and c'$\in$ C, for every w$\in$ c and for every w'$\in$ c', $[\![S]\!]_{c,w} = [\![S]\!]_{c',w'}$.

Under this assumption, which is independent of the present theoretical framework, S is seen, ideally, as a fixed content:[26] i.e. an invariant set of constraints over the admissible representations of the world in which the events in question take place, of the time (or place) at which they occur, of the teller who narrates them, and so on. Needless to say, this kind of content (which is determined by the literal meaning of the text) is not sufficient to account for the meaning of paratextual sentences such as *In S, φ*. The problem is that the interpretation of these sentences depends not only upon the literal content of S, but also upon additional assumptions about what is left implicit in S. These assumptions, concerning the suitable background against which the story should be interpreted, are part of the presumed common ground c, the current context of evaluation. So, an idealized (and simplified) account of this process is the following:

(i) When a sentence of type *In S, φ* is evaluated with respect to a context c (which can be inconsistent with S), let us make, in c, the minimal changes required to circumscribe that part of c which is consistent with S. Let c^* be this contraction of c.

(ii) The information provided by S is added to c^*. The result is the revised context c^*S.

There are several ways[27] of formalizing such a process of revision of c with respect to S, but I will not address this problem here and I will simply assume that one of these solutions is adopted and that c^*S is the intended revision of c with respect to S.

If S is a term denoting a context in the sense defined above and φ is a sentence, in the present theoretical framework the expression *In S*, in a sentence of type *In S φ*, is treated as a context-shifter. This means that, semantically, it can be interpreted as a revision function f whose argument is the current context c (i. e. the context in which the prefixed sentence is evaluated) and whose value is c^*S. As a consequence, ignoring the internal structure of the sentence φ, the truth-conditions associated with an operator such as *In S* are the following:

$[\![\text{In S } (\varphi)]\!]_{c,w}(w) = 1$ iff $[\![\varphi]\!]_{f(c),w'}(w') = 1$ for every w'∈ f(c), where $f(c) = c^*S$.

---

[26] Or, more exactly, as the fixed content which is associated with a *particular* version of the text. This specification might be necessary to avoid the objection that, for example, the first critical edition of the *Recherche* is different from the second one or that the text might have been slightly different if Proust had changed something before delivering the manuscript to the publisher. From this point of view, the idea is that the reference of a term such as *War and Peace* is contextually fixed and that, as such, it denotes the same (version of the) story in every situation.

[27] A proposal based on Lewis's system of "spheres" is presented in Grove (1988).

In other words, for any context c and for any situation w in c, the sentence *In S* φ is true in w with respect to c iff φ is true, with respect to the revised context c^*S, in every situation in c^*S. Thus, the effect of the operator *In S* is a context shift from c (the current context) to f(c), i. e. the context c^*S that we obtain if the information provided by the story S is added to the (relevant) part of the current context c which is maximally consistent with that story.

As I have just recalled, intuitively speaking a story S, in its literal meaning, is a fixed set of propositions which can be expanded in a number of ways, depending on the context (the parameter c in the above definition). The idea is that this literal meaning is a set of *constraints* over the possible interpretations of S: more exactly, it is what remains unchanged across the different interpretations of S that we get when passing from a given background of assumptions to another one. For example, whilst a sentence like *In the Recherche, Bergotte is a famous novelist* is true in any context c with respect to which we evaluate it, the truth (or falsity) of a sentence like *In the Recherche, Combray is closer to Paris than Balbec* depends on the nature of c, that is on the assumptions we might make on the "geography" of the *Recherche* (e. g. by associating Combray to Illiers and Balbec to Cabourg: an assumption which is not forced by the story itself, of course, but by independent information).

Acknowledgements

References

Arló-Costa, Horacio. 2003. Epistemological foundations for the representation of discourse context. In this volume.

Bonomi, Andrea. 1977. Indices and contexts. The ital. trans. is sect. II of Bonomi (1979); part of its content is taken up in Bonomi (1998).

Bonomi, Andrea. 1979. *Universi di discorso*. Milano: Feltrinelli.

Bonomi, Andrea. 1987. *Le immagini dei nomi*. Milano: Garzanti.

Bonomi, Andrea. 1994. *Lo spirito della narrazione*. Milano: Bompiani.

Bonomi, Andrea. 1998. Indices and contexts of discourse. *Lingua e Stile* 33:471–85.

Bonomi, Andrea and Alessandro Zucchi. 2001. *Tempo e Linguaggio*. Milano: Bruno Mondadori Editore.

Carlson, Gregory N. and Francis Jeffrey Pelletier, eds. 1995. *The Generic Book*. Chicago, IL: Chicago University Press.

Chierchia, Gennaro and Sally McConnell-Ginet. 1992. *Meaning and Grammar: An Introduction to Semantics*. Cambridge, Massachusetts: The MIT Press.

Cranston, Maurice. 1991. *The Noble Savage (Jean-Jacques Rousseau, 1754-1762)*. London: Allen Lane, The Penguin Press.

Currie, Gregory. 1990. *The Nature of Fiction*. Cambridge: Cambridge University Press.

Fauconnier, Gilles. 1985. *Mental Spaces*. Cambridge, Massachusetts: The MIT Press.

Grove, Adam. 1988. Two modelings for theory change. *Journal of Philosophical Logic* 17:157–170.

Kripke, Saul. 1973. The John Locke lectures. ms.

Lewis, David K. 1978. Truth in fiction. *American Philosophical Quarterly* 15:37–46. Reprinted in D. K. Lewis, *Philosophical Papers*, Vol. 1, Oxford: Oxford University Press, 1983.

Percus, Orin. 2002. Constraints on some other variables in syntax. *Natural Language Semantics* 8:173–229.

Prior, Arthur N. 1968. *Papers on Time and Tense*. Oxford: Clarendon Press.

Ross, Jeff. 1997. *The Semantics of Media*. Dordrecht: Kluwer Academic Publishers.

Stalnaker, Robert C. 1999. *Context and Content. Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press.

Thomasson, Amie. 1999. *Fiction and Metaphysics*. Cambridge: Cambridge University Press.

Zucchi, Sandro. 2001. Tense in fiction. In C. Cecchetto, G. Chierchia, and M. T. Guasti, eds., *Semantic Interfaces: Reference, Anaphora and Aspect*, pages 320–355. Stanford: CSLI.

March 22, 2006

11

Context and Philosophy of Science

R.A.Young

This paper aims to define a concept of *context of justification* close to Reichenbach's (section 11.1). For him, the context of justification of a theory is the context of its *rational discovery*. To prepare for a definition of context, we review the relationship between context in philosophy of language and context in philosophy of science (section 11.2). Afterwards, the relationship between context and availability of theories, methods and data is reviewed in section 11.3. In section 11.4, we sketch a more formal account of how there can be a logic of rational discovery in observational empirical science. Then, in section 11.5, we begin to extend this formal approach to experimental science. In section 11.6, we provide a definition of context and of justification within it. Next, we criticise the idea of a universal method for science that transcends all contexts (section 11.7) and the idea of a theory of everything. Finally, we explore the role of context in explanation (section 11.8) and relate different contexts of justification (section 11.9).

11.1 Introduction

This paper applies to philosophy of science concepts of *context of justification* and of *rational reconstruction* influenced by Reichenbach (1938). Moreover, it proposes a formal framework for the study of contexts of justification. Thus the paper sets out to build bridges between historically and linguistically oriented philosophy of science, formal learning theory, and formal theory of context, but the bridges involve some reconstruction of each of these fields.

In one long-standing philosophical usage, the term 'context' marks

both the context of discovery of a scientific theory and the context of its justification. Reichenbach (1938, pp. 4–7) is credited, e.g. by Nickles (2000, p. 87), with first making the distinction between these two contexts. He emphasises that philosophy of science is concerned not with the context of actual discovery, but with the context of justification. The former concerns the actual genesis of a theory and is the subject of empirical sociology. In Reichenbach's view, the latter concerns the enabling conditions for *rational* discovery of a theory. It is arrived at by rational reconstruction of the theory's genesis. It provides a rational justification for selecting the theory at the time of its discovery.

A rational reconstruction of a theory's genesis is an account of how it might have been discovered by rational use of available information. This reconstruction is historical because it works within historical constraints on the availability of information. Nevertheless, the actual historical process of discovery might be faulty. We can contrast a faulty actual historical process of discovery of a good theory with reconstructions of how it might have been rationally discovered. In the case of a discovery actually made rationally, we may be able to enhance its justification by reconstructions that show that many paths would have led rationally to the same conclusion[1]. That is to say, in the precise terms defined in section 11.6, we can show the discovery to be strongly, and not just weakly, justified. In making his distinction, Reichenbach intended to argue that rational methodology[2] of theory discovery, not the happenstance of actual historical causation, is relevant to the justification of theories.

In the second half of the twentieth century, following Popper (1959, p. 315) rather than Reichenbach, the distinction between context of discovery and context of justification was used to promote the idea that there can be no rational method of discovery. In a simple Popperian view, we are justified in holding a theory so long as it is not falsified, it is as well corroborated as any known rivals, and it has at least as much empirical content as them. However, there is no rational method of generating new theories. Of course, Popperians had opponents. Some of these opposing views have similarities with the view proposed in

---

[1]In Williamson's approach Williamson (2000), knowledge, which we have when the actual process of discovery has been rational, is not simply understood as externally true and subjectively justified, but as securely anchored in its historical environment, robust against rebuttal because there are many paths to reach evidence for it and relatively few paths that appear to give evidence against it. In my view, Williamson's approach and Reichenbach's can complement each other.

[2]Reichenbach had an account of methodology that focused on probability, and it will transpire that the methodology of this paper is developed on a different basis, but nevertheless there are similarities of overall approach.

this paper. For example, Kuhn (1970) thought that rational and determinate methods of theory generation vary with 'paradigms', but that revolutions that lead to new 'paradigms' follow no determinate method. The view proposed within this paper may also be compared with that of Lakatos (1970). In his view, there may be rational methods for theory discovery within 'research programmes', and there may also be methods for discriminating between existing 'research programmes', but there is no universal determinate method for discovery of theories. However, neither of these figures offer formal analyses, indeed perhaps they would have been antagonistic to formalisation.

Recently, Kelly (1996) has favoured a theory of rational discovery. Kelly's position will be used as one building block for the present paper, because it does offer a formal approach. His position is based on formal learning theory. This investigates whether there are reliable methods within given paradigms (Jain et al., 1999, p. 5) or background assumptions (Kelly, 1996, p. 11) for generating theories that will explain the whole of a potentially infinite data stream. The use of the term 'paradigm' suggests an analogy with Kuhn. However, in formal learning theory, a paradigm is a set of propositions that are assumed. This set of propositions may be equated with the set of models with which they are consistent, or a set of possible worlds (Kelly, 1996, p. 11). In contrast, the Kuhnian notion of a paradigm is not well-defined, but seems to include paradigmatic examples of a theory's success, which guide the work of scientists within the paradigm, as well as values and methods. Another source for the ideas of the paper, from within philosophy of science, is van Fraassen (1980), who is important both for his contextual theory of explanation and for his empiricism in philosophy of science. Some of the ideas for this paper have come from outside philosophy of science, for example from philosophy of language, and also from logic (see for example Benerecetti, Bouquet and Ghidini, Chapter 1 and Ghidini and Giunchiglia, Chapter 2 in this volume).

11.2 Context, Science and Language

To explore context in philosophy of science, it helps to begin with philosophy of language. One standard meaning of 'context' in philosophy of language is the situation in which a linguistic act occurs, as recognised in Hale and Wright (1997, p. 657). Indexicals and demonstratives, reference and the illocutionary force of speech acts all depend on context in this sense. When I use 'this television set' to refer to the tv on which I am watching the rugby match, complaining that it is not working well, and my neighbour responds to my telephone call and in-

vites me round saying 'Well this television set ain't misbehaving', then the referent is different, because I speak from number one in our Terrace, and my neighbour responds from number two. Each place and time is understood as being at some index in a possible world (the actual world is understood to be one amongst other possible worlds). Each possible world is understood to be completely determinate and as entirely describable in some metalanguage.

This conception of indexing to place and time in philosophy of language influenced philosophy of science through Putnam (1981). In Putnam's philosophy, semantic content depends upon the referents that are available at one's index. These are subject to causal constraints on reference. At this index in this world, what one means by the character string 'water' is H_2O, because at this index the available referents, which one is able to pick out, consist of liquids formed from H_2O molecules. This is what our scientists study and our scientists know water as H_2O, but, in Putnam's view, even a scientist who does not know water as H_2O, will still have semantic content about H_2O provided his available referents are H_2O molecules. Thus it is the substance which Thales (Kirk, 1957, pp. 92-3) thought to be the sole substance in the universe, because, at the index of the presocratics, H_2O molecules were the paradigm referents for Thales' term for his one substance (he used the Greek term for water). However, Putnam imagines a 'Twin Earth' in which the available referents consist of XYZ molecules, which, in the absence of sophisticated scientific tests, have the same appearance and behaviour as H_2O molecules. 'Twin Earth' might be at another index (position in space time) in this world, or at some index in some other possible world. For people on Twin Earth, the string 'water' would mean XYZ molecules. The stuff that they pick out with the string consists of XYZ molecules. Thus the difference in meaning between the character string 'water' as used by us and as used on Twin Earth is determined by the difference in available referents at different indices [3]. This variation is comparable with, although more complicated than, the variation in reference for 'this television' between myself and my neighbour.

In contrast to the accounts offered by such thinkers as Kuhn and Feyerabend (1975), Putnam provides an account according to which scientists with radically different theories about a substance may nevertheless share a term, or at least a concept, of that substance. Thus

---

[3]In order to understand the context of justification of a scientific community at time t, the present paper incorporates in its sense of 'context' available features of the surrounding region of space and time beyond what is assumed, or presupposed in the speech or thought of members of that community at time t. Thus the present sense of 'context' differs from other papers in the present volume.

Thales and contemporary scientists share the same meaning for their term for water even though Thales has no concept of a molecule and thinks water to be the only substance in the world. They can even be thought to share a concept in the sense of sharing a function that determines the same extension in this world and in other possible worlds. Moreover, it can be thought that, if Thales were to have our data, then he would accept our theory of water. In a sense, Putnam is a realist, but at the same time he does not think that there is one 'God's eye' language in which we can objectively describe the semantics of each language. For Putnam, we must describe the semantics using our own working language, and there is no objective language that can somehow stand above or beyond it. During the 1980s, Putnam described his position as 'internal realist', where 'internal' connotes 'from within a language' and not 'within the head', and the 'realist' emphasises the contribution of the world beyond the head to meaning. However, according to Putnam, our language, even though it is not a God's eye language, might be expanded to incorporate all languages. Thus, in the thought experiment, if Twin Earth is just another planet on our world, and our scientists become aware of the difference between H_2O and XYZ, then we might introduce another term 'twater' to be the translation of the Twin Earth 'water' word.

Despite Putnam, it is questionable whether Thales asserted that all the world was made of H_2O molecules, even if he did intend to assert that the world had just one substance (even this interpretation of his view is controversial, see Kirk (1957, pp. 92–3)). Of course, everyone recognises that Thales had a radically different theory from ourselves. Nevertheless, one might interpret 'water' in his theoretical discourse as rigidly designating the one substance in the world (say string[4] in the sense of string theory) instead of interpreting it as rigidly designating H_2O. Thus one might argue that translation from his language into our language is indeterminate, but perhaps Putnam would respond that this just makes Thales ambiguous. More strongly, we might argue that we cannot be confident of capturing the entire range of possible meanings. Someone with a more advanced science than our own might have an interpretation of Thales that is not expressible in our language. Even if we could add to our language terms to express some such meaning, we should not expect to express all such meanings in our language (as we shall see, in section 11.7). On the other hand, there does seem a

---

[4] "Aristotle ascribes the statement: 'Water is the material cause of all things' [to Thales] ...the views of modern theoretical physics ...are very close to the doctrines of Thales. If we replace the word water by the word string we can repeat his statement word for word from the modern point of view" Gukov (2001).

point to Putnam's idea that, over considerable historical periods, scientists may be studying the same objects in reality even though their theories of those objects vary considerably. Consider Rutherford, writing in an article for the general public (Rutherford, 1922, p. 882), when he says:

> While the negative unit of electricity exists in the form of the electron of very small mass, no evidence has been obtained that its counterpart, the positive electron of very small mass, exists. This has led to the view that the hydrogen nucleus is the *positive* electron, and that its mass is about 1,845 times that of the negative electron. This difference in mass between between the units of positive and negative electricity appears to be fundamental and offers an explanation of the asymmetrical distribution of positive and negative electricity in the structure of atoms.

Here Rutherford, speaking before the new quantum theory had reached fruition, uses the word electron in such a way that a proton might be considered a positive electron, and yet there is point in interpreting him to mean by 'electron' just what we do and to be mistaken in denying that positive particles (positrons) of the same mass exist. He does not share the theories of later scientists, but perhaps he shares some questions with later scientists and answered them falsely in 1922. After all, positrons were discovered in the next decade (Kragh, 1999) and, in the course of the discovery, it was shown that scientists had had evidence capable of corroborating their existence for years. The evidence had been misinterpreted as electrons randomly moving toward a source, when its explanation was positrons moving away from that source (Dirac, 1978, pp. 17-18).

In the present paper, the conception of context is intended to allow us to interpret Rutherford as sharing questions with contemporary scientists. At the same time, it is a conception of context which can be deployed in a fine-grained or coarse-grained way (Chapter 1 this volume), such that we may recognise a context that is shared with Rutherford, but also, at a more fine-grained level, recognise differences between the present context and that of Rutherford. Contexts will be associated with situation, in the sense of place and time, and with constraints that arise from place and time..

We do not suppose with Putnam that, if our language were sufficiently expanded then it would provide a comprehensive metalanguage for all other languages. Hence we shall not suppose, with Putnam, that there is any ultimate preferred description for the referents of our 'water' term, or for 'positive electron'. However, we shall have a place for non-observable states in our account of context and therefore we will

have an account of how the available referents may be shared in a context between scientists who do not share a theoretical understanding of what those states are.

11.3 Context and Availability

We shall think of context as something that is available to the scientist in or *from* his/her place and time. What scientists can use in communication or theorising is what is available to them. Complete descriptions of universes are not available to them. What is available to each scientist is somewhat different from what is available to others. If we think of different communities of scientists at different places and times, then there will be radical differences in what is available. We can think of 'availability' in different ways — availability in memory or availability through engaging in external activity, that is through observation, experiment and communication. Here are some ways in which information[5] is available to the scientist from his/her position, p, and time, t:

1. Sensory or memory states that are actual at t,

2. States available through sensing that may be accomplished at p after t,

3. States available from p after t through some combination of the following with sensing:

    (a) through some finite sequence of movement,
    (b) through some finite sequence of actions on the environment,
    (c) through a finite set of external instruments,
    (d) through adopting a finite set of different procedures and or axioms for reasoning,
    (e) through some finite communication in the scientist's language,
    (f) through learning a finite set of different theoretical languages with finite vocabularies and finitely specifiable (recursively axiomatisable) grammars.

If we are to define context in a way that is fruitful for understanding rational justification, then we need to consider what information we should take to be available within the context of justification. It is not fruitful to limit the relevant context to what is available in an individual scientist's memory.

---

[5] Despite its focus on information, the definition of context in sub-section (11.6.1) is not simply informational, but incorporates systems with which we interact causally, but of which we may not be capable of gaining full information.

One consideration is that we are normally concerned with rational justification in a scientific community, and in that case we need to consider whether proposed theories will prove justifiable if the community pools its information though communication. The thought is that when we adopt a theory we anticipate that it would prove justifiable if we pooled our information exhaustively. Scientists do not normally actually pool information exhaustively, but only enough salient information to check whether further pooling is likely to undermine a proposed theory. In our rational reconstruction of the context of justification, we could try to envisage complete pooling (in each scientist) of the information distributed around the memories of all the scientists in the community. However, this would be so far removed from what actually takes place in scientific communities, as to be of doubtful relevance in practise. Indeed, nowadays there is often so much information available that no individual scientist can comprehend it all. Thus, it is not clear what it would be for human scientists to pool all their information, unless what we mean by 'pooling' is simply that the information is made publicly available. Therefore we should not attempt to conceive of the context of rational justification as if it consists of complete actual cognitive pooling of information throughout the community. Instead we should think of it as including potential data streams that each scientist would receive if information were to be pooled.

A further consideration is that much of the information of a scientific community is available on paper and in electronic format. We need to consider communal resources other than human memory, when we are considering the context of rational justification. If all I need rationally to reject a theory is simply to take a book from a shelf to check a figure or to access my computer, then I am not rationally justified in holding the theory even if it is consistent with everything I hold in my memory at present. If anyone wants to reject this proposition, then they need to consider a truism about human memory. Retrieval from memory is not altogether reliable. If we took what is consciously available to a scientist at a given moment to be the context of justification, then the context would be much too narrow for justification of a scientific theory. On the other hand, if we go beyond what is consciously available and include what is available from memory, then we might just as well include what is available from other people and from libraries. With both internal memory and external sources of information, we can identify potential access to information with a branching tree of potential data streams. I propose to define context, for the purposes of scientific justification, to include a tree of potential branching data streams. For different purposes of analysis it will be possible to consider very restricted branching

trees, for other purposes the branching might ramify prolifically. These data streams may include some data streams that have their source in observation and experiment, and not just data streams from memory, communications from others, and externally stored data. Thus a rich diversity of contexts might perhaps be considered as plausible contexts of justification of theories. At one extreme there is the very restricted context in which available information consists of just what is available in the memory of one scientist at some specific time. At the other extreme, we might think that we could include in context or 'super-context' all the potential data streams that would be available were we to allow for all the novel ways in which information might be made available in science. In section 11.7, I shall argue that, if contexts are to be describable, then we cannot expect to describe such a supercontext. On the other hand, whilst the most limited context is describable, it is too restricted to provide us with an account of justification in empirical science.

Whether a scientist, in context, is justified in some claim will depend upon the restrictions, in that context, on availability of information. Thus Aristotle thought, for good empirical reasons, that the Earth was stationary, because, if the earth were moving, then parallax effects would be observable with respect to the stars (Aristotle, 1941, pp. 433-4). No such parallax effects were observable in the context of the Ancient Greeks, at least not if it is taken to exclude information through telescopes. Admittedly, if Aristotle had postulated the immense astronomical distances that we now accept, then he would have appreciated that parallax effects would be unobservable to the naked eye. However, we may take it that, in his context, the preferred models would be models with modest astronomical distances. Why should immense distances be postulated unless necessary? Thus we need not interpret Aristotle as simply ruling out immense distances by fiat, but as preferring more modest models, and, in context, never needing to abandon them. We can envisage that, in the context, there is a certain preference ordering on astronomical models (perhaps models with smaller distances are preferred to larger ones, or ones with circular motion of astronomical bodies to ones with other motions, or perhaps other considerations also matter). If the context of justification is to determine which is the right theory for Aristotle to choose, then context needs to determine a preference ordering on theories. Therefore I propose to include this in my definition of context 11.6.1.

Our interpretation of Aristotle's beliefs and preferences is constrained by the texts, but when we consider his context to exclude telescopes, we make a strategic choice. We could think of the technol-

ogy as being 'within reach' of Ancient Greece. Whether we think this or not depends upon whether we identify the context in an historically fine-grained way or in a coarse-grained way. At the beginning of this paper, Reichenbach was mentioned and it was pointed out that, for him, the context of justification is identified in rationally reconstructing how a theory could have been *rationally* discovered in the time and place where it originated. Choices about whether to be fine-grained or coarse-grained in interpreting the history are a component of rational reconstruction. There are many choices as to what we take the context of Aristotle to be. For example, we could postulate a context in which Aristotle (or his successors) can observe into the indefinite future, but only from the region of the Earth known to the Greeks of Aristotle's time, and only with the naked eye, not with telescopes.

11.4 Rational and reliable methodology

We need to find a more formal expression of the account of context proposed in section 11.3. We shall develop this by reference to Kelly's Kelly (1996) logic of reliable inquiry. Kelly's 'logic' is designed to prove reliability for methods that are applied to potentially infinite data streams. In this logic, a data stream is a denumerably infinite sequence of instances of elements from some set (D). D might be the set of natural numbers, or characters in a finite alphabet, or, more concretely, the output of some measuring instrument. Let S be a set of data streams such that some set of assumptions A holds. Let H be the set of hypotheses, consistent with A, about a way of generating each $s \in S$. A method for selecting hypotheses can be identified with a function from finite sequences of data streams to $h \in H$. A method is reliable for selecting hypotheses for some S, if for each $s \in S$, there is some finite size of sequence f such that, if an input sequence is larger than that size, then the method identifies an hypothesis which is correct for that s (the minimum size of f may vary from one data stream to another; all that is required is that the size be finite for each $s \in S$). We shall see that, given appropriate assumptions, reliability of methods is provable in nontrivial cases. Consider a human scientist applying a reliable method to some $s \in S$, the scientist will eventually hold an hypothesis that is predictive of the (denumerably infinite) s. Yet the scientist need not ever know (be able to prove) that he or she holds a correct hypothesis, because the scientist need not know the minimum f for the given data stream. The scientist, or a computer program applying the method, will never halt with the decision that some hypothesis holds. Yet the scientist who has a reliable method (or a computer program

applying the method) will converge in the limit on an hypothesis that is predictive of the entire data stream.

It is standard to consider computable methods. However one could, for example, consider an oracle machine Turing (1939) with a module for solving the halting problem for Turing machines. Kelly (1996)'s Kelly (1996) work derives from earlier work by Putnam (1965), Gold (1965) and formal learning theorists such as Osherson et al. (1986)

Consider a simple example to which we can apply formal learning theory (in substance this is an example considered by Gold (1965) in his seminal paper. Let A consist of the assumptions that

1. the set D consists of characters in a finite alphabet
2. s is such that it could be generated by some finite state machine that inputs one element of s and outputs the next.

Let H be the set of hypotheses, such that each $h \in H$ is the hypothesis that s can be generated by some specific finite state machine, and there is just one hypothesis in the set for each finite state machine. Take any $h \in H$ This hypothesis is not decidable for both a yes and no answer in finite time. Instead it is *decidable in the limit*. For a no answer, it can be decided in finite time, but for a yes answer it can only be *decided in the limit*. A procedure for *deciding in the limit* is to take an ordering of H by Gödel numbers for finite state machines, and then, taking each element h_i of H in order, to hypothesise h_i, but to retract h_i if h_i is proved false in finite time. Given assumption 2, within some finite time (variable from data stream to data stream) this procedure will eventually reach an hypothesis h_j that is true, once the method reaches such an h_j then h_j is held for all time. Thus it is not decided within finite time that h_j is true, but h_j is found within finite time, and then is never abandoned[6]

It is important to recognise that, if a data stream is such that it can be generated by one finite state machine, then it can also be generated by an infinity of others. Moreover, not all these finite state machines need be functionally equivalent (in input-output terms). Take any machine that will generate s, and call it c. There can be other machines that, given the first element of s, when they begin in a particular internal state will generate s. Yet these machines can be such that if the first data element had been different, then they would have generated

---

[6]Even though any finite state machine will eventually loop, it is not possible to decide on a yes answer after noticing that the machine loops for any finite number of times. It could just be that a finite state machine with a sufficient number of states is mimicking the behaviour of a smaller machine, whilst, in effect, counting upward through its states, ultimately coming to a state with a new effect, breaking out of the loop.

a different data stream from the one c would have generated. Thus we can think of a reliable method as generating an empirically adequate hypothesis for s but not the only one. Or we can think of a reliable method as identifying an infinite set of hypotheses that are empirically equivalent given the first element of s. Admittedly, if our Gödel numbering orders hypotheses about finite state machines from best to worse, then the hypothesis that is held to the limit will be best, but can we justify any one ordering against others?

There are many paradigms to which formal learning theory may be applied including indeterministic ones and sometimes, given the assumptions of the paradigm, there will be one uniquely right hypothesis. Jain et al. (1999) use the following indeterministic example to introduce the theory. It is assumed that one of a set of systems is generating the data stream. Each of these systems may be represented by an infinite set consisting of the natural numbers except for one missing number. In that case, the learning problem is to find the number that is missing from the actual generator stream. There are the following additional assumptions. The scientist's data stream is assumed to consist of the set of numbers in the generator in an arbitrary order. The numbers each occur at least once in the data stream. It is provable that, if the learning system has the following simple algorithm, then it will eventually have the right hypothesis about the missing number. At any given moment, the learning system hypothesises that the missing number is the lowest number missing from the data stream. If the number it is currently hypothesising occurs at time t, then it checks its record of the data stream up to time t for the lowest missing number up to that time and hypothesises that that number is the missing number.

11.5 Systems, Method and Coherence

The accounts of formal learning considered so far do not provide for experiment. The following formal specifications do seek to provide for experiment. First we define what an experimental system is for our purposes, then we define what a course of inquiry into a system is, what an hypothesis is and what it is for an hypothesis to be coherent with a course of inquiry into a system. These definitions are preliminary to the definition of context, and related definitions of justification, in section 11.6.

11.5.1 System and Global State

The definition that follows provides for non-deterministic systems that are experimentally manipulable.. At any given time, the global state of a system consists in a set of instantiations of its state variables to

values. Some of these state variables are observables. In an experimental system some of the state variables are manipulable. A system may also have hidden states that are neither observable nor manipulable. Each system has a transition relation.

A system $\mathfrak{S}$ is a tuple $\langle X, V, O, M, S \rangle$ where X is an ordered set of k state variables, V is an ordered set of k sets such that if x_i is the ith element of X and v_i of V then v_i is the set of possible values for x_i, O is the subset of observable variables of X, M is the subset of manipulable variables of X, and S is a state transition relation between states of the system. If a system $\mathfrak{S}$ has k state variables, then a global state of $\mathfrak{S}$ at a given time consists in a set $\{\langle x_0, v_{x_0} \in V_0 \rangle, \ldots, \langle x_i, v_{x_i} \in V_i \rangle, \ldots, \langle x_{k-1}, v_{x_{k-1}} \in V_{k-1} \rangle\}$ of instantiations of the state variables to values. If the transition relation holds between two global states G_t and $G_{t+\delta t}$ of the system and a time difference δt, then the global state $G_{t+\delta t}$ at $t + \delta t$ is possible given the global state G_t at t. In the special case of $S(\{\}, G, 0)$, then G is a possible global state of the system at commencement[7].

11.5.2 State, and Course, of Inquiry

Now that we have a definition of what a system is we need to define what constitutes a course of inquiry into it. This is a sequence of states of inquiry into the system, where a state of inquiry consists in instantiations to (at least some of) the observable and manipulable states of the system.

Let G be a global state of the system $\mathfrak{S} = \langle X, V, O, M, S \rangle$. In that case, $Q \subseteq G$ is a state of inquiry into the system if and only if $\forall x \forall y (\langle x, y \rangle \in Q \implies (x \in O \vee x \in M))$. A course of inquiry SQ into the system consists of a sequence of states of inquiry into the system $\{Q_0, \ldots, Q_i, \ldots, Q_j\}$ which fulfil the following condition. There is a possible sequence of global states of the system $SG = \{G_0, \ldots, G_i, \ldots, G_j\}$ such that the relation S obtains in it and $\forall i \ (Q_i \subseteq G_i)$.

In specifying a system, we have not assumed discrete time steps. Nor have we in our general specification of a course of inquiry. Nevertheless, we shall assume that all *feasible* courses of inquiry do specify discrete time steps.

---

[7]The definition of system given here is close to that of Kelly (1996, pp. 348–9), except that his definition of the transition relation does not provide for any restriction on initial states of the system, and his definition of a transition relation is envisaged as a relation between discrete steps, whereas δt in the present definition could be a time difference in the continuum. In the present paper, it is methods that impose discrete steps, whereas systems need not.

11.5.3 Hypotheses and Method

We are interested in scientists who not only observe, but who test hypotheses according to some experimental method. Therefore we need to define what an hypothesis is and when it is true. We distinguish between theoretical hypotheses, which are conjectures about what the system is that is under investigation and state hypotheses that are conjectures about the state of the system at a particular time. A method for testing a given theoretical hypothesis is a function that, given an hypothesis and a pre-existing sequence of states of inquiry determines the state hypothesis that the scientist is to hold and the state variables that he/she is to manipulate and to observe.

A theoretical hypothesis h specifies a tuple $\langle X', V', O', M', S' \rangle$. If the tuple specified by h is the same as $\mathfrak{S}$ then h is true of $\mathfrak{S}$ otherwise it is false. A state hypothesis is a specification of a pair of a global state and a time $\langle G, t \rangle$. It is true of $\mathfrak{S}$ if the global state of $\mathfrak{S}$ at t is G. A method β consists of a function from pairs of theoretical hypotheses and sequences of states of inquiry, to the set of pairs AQ which fulfils the following condition. For each $\langle a_t, MOQ_t \rangle \in AQ$, a_t is a state hypothesis and MOQ_t is a set of pairs of instantiations of state variables and times. For these instantiations in MOQ_t, the following obtains $\forall x \forall y \forall t_i (\langle \langle x, y \rangle, t_i \rangle \in MOQ_t \implies (x \in M \vee x \in O))$. For each t, MOQ_t specifies the manipulable values that are to be obtained, and the observable values that are to observed, subsequent to t. The method is followed for some theoretical hypothesis, if, for each t the state hypothesis a_t is held and MOQ_t specifies values of manipulable variables that do obtain at the time specified.

As with courses of inquiry, we shall assume that all feasible methods do specify discrete time steps.

11.5.4 Retrodiction and Coherence

A scientist may test an hypothesis in a course of inquiry by following a method. However, we need to establish just what a scientist can reasonably hope to achieve in such testing. We begin to set up the framework for our analysis in this subsection. We define retrodiction, as deduction of the possibility[8] of the course of inquiry so far. The scientist may modestly hope at least for this. Then we define coherence with the set of possible courses of inquiry. A scientist may more ambitiously hope that his/her method will guarantee not only to retrodict the actual course of inquiry until the present, but to predict the future course

---

[8]It is only a possibility that may be deduced in the case that the system is indeterministic

of inquiry and even correctly to describe all possible courses of inquiry that could arise from applying a given method to a concrete instance of a system. In section 11.6 we shall explain how it may be proved (given a relevant set of assumptions) that a method guarantees coherence, not just with the actual course of inquiry, but with all possible states of inquiry.

If $\mathfrak{S}$ is a system, β is a method and h is an hypothesis, then let $Inq(\mathfrak{S}, \beta, h)$ be the set of all the possible courses of inquiry (confirming or disconfirming of h) for $\mathfrak{S},\beta,h$.

For each $P \in Inq(\mathfrak{S}, \beta, h)$, let P_n be the course of inquiry until the nth.time step, let $c_h(a, P_n)$ be a function that evaluates to 1 if the hypothesis h is such that the possibility[9] of the course of inquiry up to n may be deduced[10] from it together with the current state hypothesis a (generated by following the method β for the nth time step), and evaluates to 0 otherwise. If this function evaluates to 1, then the course of inquiry up until now is said to be retrodicted

h is coherent with all potential courses of inquiry by β in the limit for $\mathfrak{S}$ if and only if

$$\forall p \in Inq(\mathfrak{S}, \beta, h) \exists a \exists m \forall n \geq m [c_h(a, p_n) = 1]$$

11.5.5   Theorem

An hypothesis h can be false but coherent with all possible courses of inquiry for some method β.

Proof 1: It is consistent with the definitions of system and method that for some $\mathfrak{S}$ and β there is some value of a manipulable variable such that β never instantiates that variable to that value. In that case the hypothesis h may include a state transition relation which does not obtain for that value of the variable. Thus h may be coherent with all courses of inquiry for β and yet false.

Proof 2: It is consistent with the definitions of system and method that for some $\mathfrak{S}$ and β there is some time when β does not require that an observation is made for some observable. In that case the hypothesis h may include a state transition relation which does not produce the observable value that obtains at that time. Thus h may be coherent with all courses of inquiry for β and yet false.

---

[9]For a deterministic system a description of the course of inquiry could be deduced

[10]This definition presupposes a system of deduction. The system of deduction is that of the context in which the method is applied.

11.6   Rational Justification in a Context

Now let us consider rational justification in the light of the above definition of coherence in the limit. If we are considering investigation of a system that persists indefinitely, or of a nonfinite set of systems, then, at no finite point in a course of inquiry about an hypothesis h, is there proof of the claim that h is coherent in the limit, let alone proof of the claim that the hypothesis is justified. Consider justification of the claim of coherence; the strongest kind of justification for any claim is that for which there is proof at the time of making the claim. Nevertheless, there are other senses of justification in which one's claim is justified by later events. Thus an engineer has reason to warn that a bridge will fall down but cannot prove that it will. When it does fall down, we can say that his warning turned out to *have been justified when he made it*, because, not only did he have reasons for his warning, but his prediction came true. Let us call this latter sense of justification, fallible justification, because the reason one has when one makes the truth claim does not amount to proof, justification is contingent upon what transpires after one's claim is made. We can apply this concept of fallible justification to science. Moreover, we can define it in such a way that fallible justification in a context requires that one is using a method that is provably reliable given the assumptions of that context. As a preliminary to defining justification, we will define context. In our informal discussion of the context of justification we thought that the context would include a tree of potential data streams possible for an investigator in some given place and time. We shall include this feature in our definition.

11.6.1   Context

First, we give a definition of context for the purpose of rational justification in science. Once that is given, we will give a definition of justification (actually it will transpire that we give two definitions)[11] For this purpose, we will take a context to be a tuple $\langle K, i_0, i_t, \mathfrak{S}, \beta, H, L, R \rangle$, where K is a set of background assumptions, i is the index (place and time) of commencement, i_t is the index which the observer has reached, $\mathfrak{S}$ is a system to be investigated from the index of commencement[12], β is a method of investigation, H is a preference ordering of hypotheses,

---

[11]Kelly does not give a definition of context, and the analysis of justification here is substantially different from his.

[12]In sub-section 11.5.4 $Inq(\mathfrak{S}, \beta, h)$ was defined as the set of courses of inquiry for some abstract system, method and theoretical hypothesis. Some contexts may lack some variations of initial state of the abstract system $\mathfrak{S}$. Therefore only a subset of $Inq(\mathfrak{S}, \beta, h)$ may be available in a given context.

L is a language, and R is a set of rules of inference. For each $h \in H_g$ (where H_g is the ground set of the ordering H), there will be courses of inquiry $Inq(\mathfrak{S}, \beta, h)$ determined from the commencement of the context. The language L is an interpreted language with the states of $\mathfrak{S}$ as its domain and the assumptions K are expressible in L as axioms to which the rules R may be applied. A potential course of inquiry in a context proceeds from the moment of commencement and may extend indefinitely in time and space from the moment of commencement.

For our purposes, a system $\mathfrak{S}$ is a type of system, but there will be contexts of investigation (e.g. cosmology) where there is only one instance of a system available for investigation. When we speak of Rutherford and a contemporary scientist sharing a context, then we are speaking of a type of context. We can identify types of context by defining relations on contexts, such that contexts C_1 and C_2 count as being of the same type because relations hold between elements of them. For example, all the elements in the two tuples might be the same except for the third element of each, thus we would have $i_{t_1} \neq i_{t_2}$. In this type, all assumptions, methods and so on would be the same except for the place and time of current investigation. This would be suitable for the analysis of a single scientist, or group of scientists rigorously pursuing an unchanging research programme. However, we might want to recognise two research groups as being in the same type of context even though their work began independently, and thus we would have $i_{0_1} \neq i_{0_2}$. In the case of Rutherford and contemporary scientists we might recognise identity of type of context on the basis of some overlap of assumptions and methods, whilst appreciating differences. Thus the definition of context just given, together with a set of type definitions can provide for the variety of contexts envisaged in earlier sections of this paper, and may be used to analyse the approaches of Reichenbach to context, Kuhn to paradigms and Lakatos to research programmes.

11.6.2 Weak justification

Weak justification of a theoretical hypothesis obtains when the hypothesis is consistent with, and explanatory of, the entire actual data stream (including the future) for some instance of a system. Justification is only weak, because the hypothesis need not be compatible with all potential data streams.

Let $h \in H$ be an hypothesis that is being investigated for the system $\mathfrak{S}$ in the context $\mathfrak{C} = \langle K, i_0, i_t, \mathfrak{S}, \beta, H, L, R \rangle$ using method β. Let $p \in Inq(\mathfrak{S}, \beta, h)$ be the actual course of inquiry.

An hypothesis h is weakly fallibly justified for the system $\mathfrak{S}$ in

context $\mathfrak{C}$ with course of inquiry p if and only if

- $\exists n < t \forall m \geq n[c_h(p_m) = 1]$
- it is provable in L using R and assuming K that β is a reliable method for discovering an hypothesis such that $\exists n \forall m > n(c_h(p_m) = 1))$
- $\forall m > t(c_h(p_m) = 1)$
- there is no $h' \in H$ that fulfils the previous conditions and is of higher preference than h.

11.6.3 Strong Justification

In some contexts strong justification is possible for a theoretical hypothesis. This is where, given the assumptions K of a context it is provable that if the theoretical hypothesis is consistent with and explanatory of the data streams available within the context then it is coherent with all potential data streams.

An hypothesis h is strongly fallibly justified for the system $\mathfrak{S}$ in context $\mathfrak{C}$ with course of inquiry p if and only if

- h is weakly fallibly justified
- it is provable in L using R assuming K that, if h is weakly fallibly justified, then h is coherent with all potential data streams by β in the limit for $\mathfrak{S}$

Is it possible to show that, given the the background assumptions of some contexts, there are methods that enable strong justification for some non-trivial hypotheses about systems with hidden variables? Consider the two following principles of plenitude[13].

1. as the number of instances of state s for a specific instance of $\mathfrak{S}$ (in a course of inquiry conducted according to the method of the context) increases towards infinity (either by nature or through manipulation) and if $S(s, r, \delta t)$ holds. then the number of instances of r following δt afterwards also eventually increases towards infinity

2. for each initial configuration which is possible for $\mathfrak{S}$, as the number of instances of $\mathfrak{S}$ in the context , is increased toward infinity, the number of instances of that initial configuration also eventually increases towards infinity.

---

[13]These two principles taken together are a variation on Kelly (1996), the main difference being that he does not make the second principle explicit, perhaps because he is considering just one instance of a system. In many scientific contexts, it seems best to consider repeated experiments with different instances of a system.

Given these two principles of plenitude as assumptions of a context, strong justification can be established. Of course, strong justification is still not enough for us to establish the truth of the theoretical hypothesis that is strongly justified, because the method need not involve manipulating all instances of manipulables or observing all instances of observables. For example, the method may consist in a sequence of discrete samplings from a continuum, thus neglecting an infinity of instantiations of state variables.

11.7 Contexts, Epistemology and Ontology

Two definitions of justification have been given, both of which are contextual. They depend upon proofs of reliability of method, and these proofs are only possible given the background assumptions, and preference orderings, available in a context. These background assumptions and preference orderings can vary, and methods vary with them, so there can be no proof of a universal method. Thus there is no universal context and no reliable method that transcends contexts. Or is this conclusion too swift? When arguments are presented to the effect that all science is contextual, a common response is that the arguments are epistemological not ontological. It may be argued that the most arguments about context can succeed in showing is that all *justifiable knowledge claims* are contextual, they cannot show that the *content of science* (the *propositions* that science depicts to be true in the world) are contextual.

Could the *content* of science itself be context dependent? It can be argued that it is. We shall argue that there is a constraint on the capacity of a scientific community to have contents. This constraint is that the capacity of any one language, even if expanded with any feasible extensions to its vocabulary, is limited in comparison with the propositions instantiated in the world. Therefore we cannot expect any one scientific community to have a comprehensive set of sciences. At its simplest, we might think of the scientific content that a community can have as a list of propositions, or as a body of propositions that may be axiomatised, or at least articulated in some language. A fundamental constraint is the *number* of propositions that can be combined together in any one human form of organisation, such as a list, or represented in a human language. If we take the number of items that may be listed, or the number of proofs that may be generated from a recursive axiomatisation and a finite set of rules, or the number of sentences that can be constructed in a language with a finite alphabet and finite (but infinitely various) sentence length, then in each case the

number of items that can be organised in this way is a denumerable infinity Boolos and Jeffrey (1980). Now, in some sense, we can conceive of larger infinities than a denumerable infinity. For example the number of points in a spatial or temporal continuum is greater than a denumerable infinity. Therefore a fundamental question about any empirical science is how the number of the propositions that might be truly asserted is related to the number of sentences available in any one language to assert them, or if you like to the number of propositions we can list, or to the number of theorems that can be proved in a system. Can we rule out the possibility that the number of propositions in the world outnumbers the number of sentences that we can form to express them? This is an ontological question not simply an epistemological question.

It may help us discuss this question, if we consider the following passage from Lewis (1984), because it attempts to address an ontological question about the relationship between language and the world. Lewis is objecting to Putnam's model theoretic argument. In the original version of his argument Putnam (1980), Putnam argued, from the Löwenheim Skolem theorem, that, if there is at least a denumerable infinity of objects in the world, then any first-order theory will be true *on some interpretation* of the language in which we can express the theory.

> When we limit ourselves to the eligible interpretations, the ones that respect the objective joints in nature, there is no longer any guarantee that (almost) any world can satisfy (almost) any theory ... eligibility ... [is] a matter of degree. The mereological sum of the coffee in my cup, the ink in this sentence, a nearby sparrow, and my left shoe is a miscellaneous mess of an object, yet its boundaries are by no means unrelated to the joints in nature. It is an eligible referent, but less eligible than some others. (I have just referred to it.) Likewise the metal things are less of an elite, eligible class than the silver things, and the green things are worse, ... but all these classes belong to the elite compared to the countless utterly miscellaneous classes of thing that there are. *Ceteris paribus*, an eligible interpretation is one that maximises the eligibility of referents overall. Yet it may assign some fairly poor referents if there is good reason to.

Lewis is arguing that ontology constrains the interpretation of language. It constrains the objects to which we can refer, and the classes that are eligible for association with predicates; they are ones whose boundaries relate to 'objective joints in nature'. These classes may relate to objective joints in a convoluted way. Thus we can refer to the set of 'utterly miscellaneous classes', even though we cannot associate

a predicate with each of them. We cannot do the latter because each of them is utterly random, and also because there is a countless number of them. Even if the number of objects in the universe is only a denumerable infinity, the number of sets (the powerset) of those objects constitutes a nondenumerable infinityBoolos and Jeffrey (1980). Within this powerset, there will be a nondenumerable infinity of utterly miscellaneous classes (as Lewis recognises), but there may also be a nondenumerable infinity of natural classes. For the sake of argument let us concede Lewis' claim that there are objective joints in nature. If we make this realist assumption, but recognise that a nondenumerable infinity of classes may *respect* these joints, then it will be impossible, in any one language, to associate predicates with all of them. Thus we have a *realist* basis for each scientific or linguistic community failing to grasp all the scientific content that might be grasped. Thus, there is an ontological argument, indeed a realist argument, for the conclusion that science is contextual. Note that the arguments just presented for the contextuality of science holds irrespective of whether the science is presented as a first order theory or as second order theory. Some second order theories require a nondenumerable domain. Nevertheless, if a second theory can be articulated in a humanly intelligible language, that language will be limited to a denumerable infinity of propositions and a denumerable infinity of predicates. Thus the model for such a language may need to be non-denumerably infinite, but the language will associate predicates with far fewer sets than there are eligible sets of objects in the domain.

I have put this ontological argument linguistically and therefore it might seem that it could successfully be challenged by proponents of a semantic analysis of theories of science. On the semantic account of what a theory is, it is a model, or class of models van Fraassen (1980). It is a mathematical structure which can be used to provide an interpretation that will make our scientific statements true. Since we can have nondenumerably infinite models, it might seem, at first sight, as if a theory, on the semantic view, could capture more than what a language can express. However, human theories must be humanly graspable. Consider a theory that is equivalent to a class of models including ones which, in all their detail, are beyond the power of any one human language to express. This humanly graspable theory, equivalent to a class of models, is an abstraction that only captures an aspect of the models in all their richness. Other humanly graspable theories might express different aspects. Even with respect to possibilities there might be modal indeterminacy as proposed by Humberstone Humberstone (1981), Young (1999).

We have just explored what I called a realist argument for science being contextual. Within philosophy of science, there has been a vigorous debate between realists and non-realists. How does the position articulated here relate to the wider debate? In order to categorise it, we need to relate it to van Fraassen's van Fraassen (1980) position, because he is the most important contemporary spokesman for empiricist non-realism in philosophy of science. The realism that he opposes declares that the aim of science is to discover literally true theories, which are true even about the non-observables. In contrast, for him, the aim of science is only to construct theories that are empirically adequate. The realist thinks that science aims to construct models that are entirely instantiated in the world. For van Fraassen, the non realist (he would say "constructive empiricist") aims to construct models whose observables are instantiated in the world.

On van Fraassen's account, there is no difference between the formal semantics of the realist and the non-realist. Thus van Fraassen's position is not like that of Dummett (1978), where non-realism goes with a semantics of warranted assertability as opposed to a semantics based upon the correspondence theory of truth. Van Fraassen's position on semantics may be compared with the account of semantics provided by one school of thought on context, namely that of Ghidini and Giunchiglia, Chapter 1, and Benerecetti, Bouquet and Ghidini, Chapter 2, this volume), who think of contexts as consisting of sets of local models. Each agent has a different set of local models, and there is no assumption of one shared model which is *the* fundamental model that is realised in actuality. Like van Fraassen, they do not develop a semantics of warranted assertability in order to express non realism. Both they and van Fraassen express their non-realism by arguing that it is unnecessary to commit to any one model.

The position of the present paper is very close to van Fraassen[14] with respect to rational aims of science, because justification has been defined in terms of coherence with the tree of potential data streams (strong justification) or the actual data stream (weak justification), not directly in term of the model being instantiated in reality. Even in the case of strong justification of an hypothesis in a context C_1, there can be coherence with all the potential data streams in C_1, and yet it might

---

[14]However, van Fraassen develops his empiricism by reference to non-instrumental 'naked eye' observation. In the present paper, no such constraint has been imposed — some 'observables', for present purposes, may only be identifiable instrumentally. Also, what constitutes an observable or manipulable is context relative. Thus van Fraassen's approach is in one way more restrictive on observables, but perhaps, on the other hand, is not not as contextual as the present approach.

turn out that, in a context C_2 with more powerful methods and thus a richer set of data streams, the hypothesis is incompatible with observation and thus false. Despite general agreement with van Fraassen, I have argued in the present section that even if the realist chanced to have a class of models one of which was instantiated in reality, this would not prevent the content of science from being contextual. An innumerable number of other realists might have chanced on interestingly different classes of models, also instanced in reality.

11.8   Explanation and Context

One argument that the realist has against van Fraassen is that genuine explanations must describe real causes. Therefore, if the scientist aims at explanation, the scientist must aim at more than mere empirical adequacy, in other words at literal truth. Van Fraassen's response is that explanations need not be true. We might attempt to side-step this issue, on the grounds that, according to the argument of the previous section, what is important for contextuality is not whether our models are realised in reality, but whether our grasp of them can be a grasp of the whole reality, for example of the entire cause of something. Nevertheless, it is important to explore the question of how explanation fits into a theory of context in science, and since van Fraassen offers a contextual account of explanation, the issue arises of whether his account can provide us with the account of explanation that we need.

Van Fraassen's account of explanation is a particular kind of contextual account, because it conceives of explanations as answers to questions. Thus van Fraassen says (van Fraassen, 1980, p. 134)

'I shall now propose a new theory of explanation. An explanation is not the same as a proposition, or an argument, or list of propositions; it is an answer. An explanation is an answer to a why-question. So, a theory of explanation must be a theory of why-questions.'

Van Fraassen argues that context is required in order to specify what is requested in asking a question (van Fraassen, 1980, p. 156).

'The discussion of explanation went wrong at the very beginning when explanation was conceived of as a relationship like description: a relation between theory and fact. Really it is a three-term relation, between theory, fact and context . . . Since an explanation is an answer, it is evaluated vis-a-vis a question, which is a request for information. But exactly what is requested, by means of the question 'Why is it the case that P?', differs from context to context. . . . to say that a given theory can be used to explain a certain fact, is always elliptic for: there is a proposition which is a telling answer, relative to this theory, to the request for information about certain facts (those counted as relevant

> for this question) that bears on a comparison between this fact which is the case and certain (contextually specified) alternatives which are not the case.'

For van Fraassen the 'why' in the request for explanation is governed by a relevance relation and therefore an 'abstract question' may be, 'at least in a preliminary way', identified with a triple $\langle P_k, X, R \rangle$, where P_k is the topic (what is to be explained), X is the contrast class $\{P_1, \ldots, P_k, \ldots\}$ and R is the relevance relation (van Fraassen, 1980, p. 143). His position on the the general structure of why questions is that they are of the following form (van Fraassen, 1980, p. 127).

Why (is it the case that) P in contrast to (other members of) X?

According to van Fraassen, both the contrast class, X, and the relevance relation R are determined by context. According to the definition given in section 11.6 of the present paper, a token context is a tuple $\langle K, i_0, i_t, \mathfrak{S}, \beta, H, L, R \rangle$, and the R in the context tuple stands for a set of rules of inference, not for a relevance relation. Nevertheless, relevance relations may be determined by context in our sense. Together the K (background assumptions), and the $h \in H_g$ given at the time of the question by β (method) and H (preference ordering on hypotheses) determine what is relevant at a particular time. Moreover, if the rules of inference include rules of inference for an erotetic logic(Young, 2001), then these may determine questions that are open given the hypothesis and the data stream between i_0 and i_t.

What constitutes an answer to a 'why' question? Van Fraassen gives the following definition of a direct answer (1980, p.144).

> B is a *direct answer* to question $Q = \langle P_k, X, R \rangle$ exactly if there is some proposition A such that A bears relation R to $\langle P_k, X \rangle$ and B is the proposition which is true exactly if P_k; *and* for all $i \neq k$; *not* P_i; and A is true.

Of course, according to van Fraassen, when a non-realist gives or accepts an answer to such a question, where the answer describes non observables, then the non-realist is not committed to the literal truth of the answer, but only to its empirical adequacy.

The set of direct answers to a question is called the set of *alternatives*. The context needs to determine this set, and the contrast class as well as the relevance relation which we have already considered. Context in our sense can determine contrast classes because K and H together with L and R, will determine a tree of contrasting potential data streams within the context (NB this tree will be different from any $Inq(\mathfrak{S}, \beta, h)$, because it will be the data streams that the scientist takes to be open rather than the data streams that the system $\mathfrak{S}$ de-

termines). It can also determine a set of alternatives, because K, H, L and R together with the data stream between i_0 and i_t, will determine contrasting hypotheses within the context.

11.9   Explanations and illocutionary acts

In the early sections of this paper, scientific communities were discussed. However, in the last three sections little has been said about different scientists interacting with each other. Let us now consider how explanation may be given by one scientist to another. On van Fraassen's account, an explanation, is an answer to a question, and a question is a request for information. A request for information is an illocutionary act (Austin, 1975, p. 162). So, to be precise, we need to distinguish between the illocutionary act of questioning, and what van Fraassen's calls the 'abstract question', which, on his view, is the triple $\langle P_k, X, R \rangle$. Indeed, perhaps we should not follow van Fraassen in using the term 'question' for both the illocutionary act and the 'abstract question'. Perhaps the illocutionary act is simply that of 'asking' a question.

If we say this then we follow Austin, who has 'ask' in his list of expositives (Austin, 1975, p. 162), together with 'explain' (Austin, 1975, p. 163). Thus both the asking of questions and the giving of explanations are in his category of clarification of reasons, arguments and communications. If the asking of a question is, at least when it is felicitous, a step in clarification, then it need not be a request for information. The asking of a question might, at least on some occasions, be more like a step in a deductive proof. In a deductive proof, all the information necessary to prove a theorem is already present, nevertheless it requires a proof to make clear that it is a theorem. On Austin's view the giving of an explanation is also, at least when felicitous, a step in clarification. As an illocutionary act, it is one that can be performed by uttering a sentence, where that sentence, in virtue of certain conventions or rules, and the surrounding context, constitutes an act. Sometimes, of course, one can perform conventional acts in a non-linguistic way. Nevertheless, clarification would seem to be a linguistic matter. If this is right, then how linguistic is an explanation itself?

Let us go back to van Fraassen. In defining the class of direct answers he defines them in terms of propositions, that is in a way that abstracts from language. Thus part of his definition is 'there is some proposition A such that A bears relation R to $\langle P_k, X \rangle$' (see p. 272). On the other hand, when he explains the role of a context he tells us that, 'The context will generally select the proposition expressed by a sentence A via a selection of referents for the terms, extensions for the predicates and

functions for the functors' (van Fraassen, 1980, p. 137). Thus at least part of the role of a context is to select the proposition expressed by a sentence. Furthermore he argues, as we have seen, that an explanation is a three part relation between theory, fact and context. On his view, theories and facts are not directly linguistic entities (theories are classes of models and facts feature in models). Nevertheless, since the role of context is at least in part linguistic, it would seem that the relation of explanation is in part a linguistic. If one thinks in a formally linguistic way about it, an explanation has only been given when a sentence has been identified as a premiss from which it is possible to deduce a description of the right element in the contrast class.

Is it then a condition of an explanation having been given that the recipient is able formally to deduce some description of what is to be explained? Or is it at least a condition of the felicitous giving of explanations that the recipient be able to make such a deduction? It would be too much to require that all scientific explanations consist of formal derivations, but is it too much to require that a scientific explanation be given so articulately that its recipients should be able to construct a formal derivation from it if necessary? I think it is not too much. Therefore we need to think of explanations not just contextually, but linguistically. Thus, I agree with van Fraassen that explanation is not a two-place relation between theory and fact but involves context as well. In my definition of context, I have already included a place for language. To allow a place for illocutionary acts, one might consider extending the set of rules to include other rules than rules of inference.

In a full account of the illocutionary act of giving a scientific explanation, we would need to ask whether the explanatory theory that the recipient of the explanation accepts needs to be the same as that employed by the person giving the explanation. I am inclined to think that this need not be so. The two people may have different models in mind, yet, so long as there are appropriate relations between the two models, the person giving the explanation may have an adequate explanation, and so may the person receiving it. Indeed, where we have different scientific communities, it is often the case that they employ different models, and that here we should think of them as having different contexts. Yet often one community may offer another community an explanation which needs to be transposed from the model of the exporting community into the model employed by the importing community. Thus aspects of quantum physics may explain aspects of chemical bonding, aspects of chemical bonding may explain aspects of biochemistry, and aspects of biochemistry may explain aspects of cellular biology, but, despite all that, we should not expect the same model

to be employed in all these sciences [15].

11.10   Conclusion

This paper set out to apply to philosophy of science a concept of 'context of justification' and of 'rational reconstruction' influenced by Reichenbach. Moreover, it proposed to provide a formal framework for the study of contexts of justification, which it offered in sections 11.5 and 11.6, after some preparation in section 11.4. In moving from the original intention to the formal framework, I offered some less formal comments on philosophy of science in sections 11.1, 11.2 and 11.3, suggesting that the framework might accommodate some points from Kuhn (1970), Lakatos (1970) and Putnam (1981), whilst indicating that the framework was not intended merely to articulate any of these positions. In the relatively formal sections, the paper was influenced by the work of citetkelly:1996. Thus the paper sought to build a bridge between historically and linguistically oriented philosophy of science and formal learning theory, but the bridge involved some reconstruction on both sides of the divide. In section 11.7, I argued that context in science is not simply an epistemological phenomenon, but pertains to limitations on the expressive power of languages to articulate properties in the world. In sections 11.8 and 11.8, I related the formal framework to van Fraassen's van Fraassen (1980) contextual account of explanation and briefly discussed an account of explanation as illocutionary. I also proposed a line of research using some logic from the Trento group(see Chapters 1 and 2) in order to analyse how explanations may be given by scientists in one context to those in another.

References

Aristotle. 1941. *De Caelo*. In R. McKeon, ed., *The Basic Works of Aristotle*. Random House. transl. Stocks, J.L.

Austin, J.L. 1975. *How to do things with words*. Oxford University Press, 2nd edn.

Boolos, G.S. and R.C. Jeffrey. 1980. *Computability and Logic*. Cambridge University Press.

Dirac, P.A.M. 1978. The development of quantum mechanics. In P. Dirac, Hora.H, and J. Shepanski, eds., *New Directions in Physics*. Wiley.

Dummett, M. 1978. *Truth and other Enigmas*. Duckworth.

Feyerabend, P.K. 1975. *Against Method*. Verso.

---

[15]To analyse the relationship between different communities, I would propose to use a version of the 'Distributed First Order Logic' under development by Ghidini and Serafini (2000)

Ghidini, C. and L. Serafini. 2000. Distributed first order logics. In D. Gabbay and M. 'De Rijke', eds., *Frontiers of Combining Systems*. Research Studies Press.

Gold. 1965. Limiting recursion. *Journal of Symbolic Logic* .

Gukov, S. 2001. Research description. Clay Institute Prize Research Fellow Statement:
http://www.claymath.org/researchers/longtermprizefellows/
gukov_research.pdf. Cambridge, Massachusetts.

Hale, B. and C.G Wright, eds. 1997. *A Companion to the Philosophy of Language*. Blackwell.

Humberstone, I.L. 1981. From worlds to possibilities. *Journal of Philosophical Logic* .

Jain, S, D.N. Osherson, J. Royer, and A. Sharma. 1999. *Systems that Learn*. MIT Press, note that this is in effect, Osherson(1986) 2nd. edn.

Kelly, K. 1996. *The Logic of Reliable Inquiry*. Oxford University Press.

Kirk, J.E., G.S.and Raven. 1957. *The Presocratic Philosophers*. Cambridge University Press, 1st edn.

Kragh, H. 1999. *Quantum Generations*. Princeton University Press.

Kuhn, T.S. 1970. *The Structure of Scientific Revolutions*. Harvard University Press, 2nd edn.

Lakatos, I. 1970. Falsification and the methodology of scientific research programmmes. In A. Lakatos, I.and Musgrave, ed., *Criticism and the Growth of Knowledge*, pages 91–195. Cambridge University Press.

Lewis, D.K. 1984. Putnam's paradox. *Australasian Journal of Philosophy* pages 221–236.

Nickles, T. 2000. Discovery. In W. Newton-Smith, ed., *Companion to the Philosophy of Science*, pages 85–95. Blackwell.

Osherson, D.N., M. Stob, and S. Weinstein. 1986. *Systems that Learn*. MIT Press.

Popper, K.R. 1959. *The Logic of Scientific Discovery*. New York: Basic Books. trans. of *Logik der Forschung*, 1934.

Putnam, H. 1965. Trial and error .... *Journal of Symbolic Logic* 30(1):49–57.

Putnam, H. 1980. Models and reality. *Journal of Symbolic Logic* pages 464–482.

Putnam, H. 1981. *Reason, Truth and History*. Cambridge University Press.

Reichenbach, H. 1938. *Experience and Prediction*. University of Chicago Press.

Rutherford, E. 1922. Matter,constitution of. In *Encyclopedia Britannica*, vol. XXXI, pages 880–883. [The Encyclopaedia Britannica Company,Ltd.], 12th edn.

Turing, A.M. 1939. Systems of logic based on ordinals. *Proceedings of the London Mathematical Society* 45:161–228. ser. 2.

van Fraassen, B.C. 1980. *The Scientific Image*. Oxford University Press.

Williamson, T. 2000. *Knowledge and Its Limits*. Oxford University Press.

Young, R.A. 1999. Context and supercontext. In P. Bouquet and L. Serafini, eds., *Proc. of the 2nd. Int. Conf. on Context*. Springer Verlag.

Young, R.A. 2001. Explanation as contextual. In V. Akman, P. Bouquet, R. Thomason, and R. Young, eds., *Modeling and Using Context: the procedings of the third international conference CONTEXT 2001*, pages 381–395. Springer Verlag.

March 22, 2006

Index